Genes of the Ovine Major Histocompatibility Complex Class II Region

Harry Wright
Moredun Research Institute
Edinburgh, Scotland.

A thesis submitted for the degree of
Doctor of Philosophy
University of Edinburgh
1995

Declaration

The work described in this thesis was carried out in the Division of Immunobiology at the Moredun Research Institute, Edinburgh. The design and execution of the experimental work and the interpretation of the results was carried out by the author. Contributions to the work of this thesis made by colleagues is fully acknowledged in the text.

Harry Wright

Abstract

The major histocompatibility complex (MHC) is a multi-gene family encoding proteins which play important roles in the immune response to antigenic challenge. Three distinct regions designated class I, II, and III have been defined in the MHC of mouse and man. This thesis focuses on the genes of the class II region of the sheep. The products of the class II alpha (*A*) and beta (*B*) genes are heterodimeric glycoproteins whose physiological function is to present exogenous peptides to helper T cells. The recognition of MHC class II/peptide complexes by the T cell receptor signals the release of a cytokine cascade resulting in T and B cell proliferation, macrophage activation and B cell differentiation with the production of increased amounts of pathogen-specific antibody.

Much is known about the detailed structure and function of the MHC of man and mouse. However, when this project began little was known about the detailed structure of the MHC of the ungulates, the economically important group of animals which contains cattle, pigs, horses and sheep. As part of a study investigating fundamental cellular immunology in the sheep, this thesis describes the characterisation and expression of the genes of the sheep MHC class II region.

Cosmid libraries prepared from DNA from three unrelated sheep were screened with probes from the *DP*, *DQ* and *DR* sub-regions of the human and mouse MHC class II regions. Cosmids were used because they facilitate the cloning of relatively large genomic inserts. Restriction maps of the cosmids have been produced showing that some of the clones overlapped. The MHC *A* and *B* genes within the clones have been sequenced and assigned to a specific sub-type. Functional genes have been identified by the reaction of their products with anti-sheep class II monoclonal antibodies following DNA-mediated transfection into the mouse L cell, a fibroblast cell line which does not express endogenous mouse class II genes. Transcription of some of the genes has been demonstrated by Northern blots and reverse transcription polymerase chain reaction.

A restriction map of the sheep class ll *DQ* sub-region has been constructed and shown to contain two distinct *DQA* loci with associated *DQB* genes. The *DQ1 A/B* gene pair was expressed in the mouse L cell. The sheep class ll *DQ1* product at the cell surface reacted with a sub-set of the available anti-sheep class ll monoclonal antibodies. The *DQ2* genes were transcribed and some evidence for their cell surface expression was obtained, although this was not formally proved.

A previous study demonstrated the expression of a putative sheep *DRA* gene when co-transfected with a sequenced *DRB* gene. The sequence of the sheep *DRA* gene is described here together with sequence data from a number of *DRB* genes or pseudogenes which show that, depending on haplotype, the sheep *DR* sub-region may contain up to five *DRB* genes.

The cloning and sequencing of ruminant orthologues of the *HLA-DNA* and *-DOB* genes is described for the first time. Although evidence was obtained for the transcription of the sheep *DNA* gene, its *DOB* gene partner is transcriptionally silent.

A class ll locus designated *DY* which is not found in mouse or man is described from the sheep MHC. Cloning and sequencing has shown that it contains a class ll *A/B* gene pair like that of the expressed *DQ1* locus, however, it appears to be transcriptionally silent.

Acknowledgements

I should like to thank a number of people for their help at various stages of this project.

First of all Dr Colin McInnes for his help and sound advice in many discussions over the past few years. Jim Redmond for his help with many aspects of the work. Mr E.V. Deverson, Prof. J.C. Howard, Prof. H.R.P. Miller, Dr G.D.Baird and Prof. I. McConnell for their help and encouragement in the early stages of this project. Dr D. Haig and Dr J. Hopkins for their constructive advice on the content of the thesis and for reading the thesis. Brian Easter for the photographic work and Diane Donaldson for her help with information retrieval.

Over the past few years I have been fortunate enough to work in lab 115 with a group of people whose company I have enjoyed very much. To Colin, Jim, Ann, Mary, Nyree, Alex, Andrea, Campbell, Keith, Jane and Vicki, many thanks.

I should also like to fully acknowledge here the contribution made by my colleague Dr Keith Ballingall to the gene transfection and FACS analysis described in this thesis.

The work was supported by Core funding from the Scottish Office Agriculture and Fisheries Department.

Table of Contents

Abbreviations

BPB	Bromophenol Blue
CIP	Calf intestinal alkaline phosphatase
DIG	Digoxygenin
DMEM	Dulbecco's modified Eagle's medium
FACS	Fluorescence activated cell sorter
FACscan	Fluorescence activated cell scanner
FITC	Fluorescein isothiocyanate
HAT	Hypoxanthine/Aminopterin/Thymidine
HEPES	(N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid])
IPA	Isopropanol
IPTG	Isopropyl-b-D-thiogalactoside
kbp	kilobase pairs
kD	kilodaltons
LB	L-broth.
mAb	Monoclonal antibody
MHC	Major Histocompatibility Complex
MLR	Mixed lymphocyte reaction
MOPS	(3-[N-Morpholino]propanesulphonic acid)
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
RFLP	Restriction fragment length polymorphism
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SDS	Sodium dodecyl sulphate
SSC	Sodium chloride/sodium citrate
TAE	Tris/Acetate/EDTA
TCA	Trichloroacetic acid
TE	10mM Tris/0.1mM EDTA
TEMED	N,N,N,N-tetramethylethylenediamine
Tk	Thymidine kinase
Tris	tris(hydroxymethyl)aminomethane
X-Gal	5-bromo-4-chloro-3-indolyl-b-D-galactoside

Chapter 1

Introduction

1.1 Historical overview of the MHC

The years since 1975 have seen the cloning and sequencing of many hundreds of genes from many and diverse species. One multi-gene family stands out as unique in terms of the extreme polymorphism of its individual gene loci. It has been found in all vertebrate species studied and its gene products play a crucial role in the immune response to infectious agents, neoplastic transformation and self-nonself discrimination. It is called the Major Histocompatibility Complex (MHC).

The first section of this introduction describes the history of the discovery of the MHC. The second section goes on to deal with the structure of its genes and proteins using the well-documented data from mouse and man, while the third describes the relationship between structure and function in the MHC. In very large measure, our understanding of the MHC comes from studies in mouse and man. However, this thesis deals with the MHC of the sheep, a member of the distinct and economically important taxonomic group, the ungulates, which also includes cattle, pigs, goats and horses. The last section of this introduction therefore describes studies of the sheep MHC up to the point at which the current work started, with reference where appropriate to the MHC of cattle which has received much more attention than that of the sheep.

In his book 'Natural History of the Major Histocompatibility Complex', Klein (1986), identified three developments which facilitated and led to the discovery of the MHC. The first was the development of inbred strains of mice, the second was the study of the phenomena associated with transplantation of tumours between mice, and the third was the study of blood-group inheritance. The road to the MHC started with observations of tissue incompatibility in mice. A mouse tumour could be transplanted between members of an inbred population of mice but could not be transplanted to wild mice. By crossing mice in which the tumour grew, with mice which were resistant to the tumour, Little and Tyzzer (1916) found that while the F1 generation was susceptible, only 1% of the F2 generation produced by backcrossing were

susceptible. This indicated the involvement of as many as fifteen genes in the phenomenon of tumour transplantation. That the same genes were involved in the transplantation of normal tissue was shown by repeating the experiment with normal spleen (Little and Jonsson, 1922).

Between 1900 and 1930, the blood group antigens, A, B and O were discovered and haemagglutination methods were described for their identification. J.B.S. Haldane made the conceptual leap that the rejection of tumour transplants was analogous to the destruction of incompatible red blood cells (RBCs) following transfusion. In other words there were tumour alloantigens analogous to the blood group antigens. Gorer undertook to prove this and immunised rabbits with RBC's from three strains of inbred mice. He used a haemagglutination assay to demonstrate the presence of four blood group antigens I-IV whose expression varied among the three strains (Gorer, 1936). When a mouse from one of the strains developed a tumour, he used it to inject mice from other strains and showed by the frequency of susceptible and resistant mice in the F1 and F2 generations, that the susceptibility was controlled by two genetic loci. Furthermore, he found that those mice which were negative for the antigen designated II, rejected the tumour and that the proportion of resistant and susceptible animals in the F1 and F2 generations of informative crosses and backcrosses, indicated that one of the two loci which controlled the growth of the tumour transplant also coded for antigen II. Moreover, antigen II was shared by normal and malignant tissue (Gorer, 1937, 1938).

The next development stemmed from the collaboration between Gorer and Snell. Snell had been producing congenic strains of mice by crossing two inbred strains that different by perhaps as many as fifteen of the tumour resistance loci. The offspring were back-crossed to one of the parental strains and this procedure was repeated over many generations. At each stage the offspring were selected for resistance to the tumour. In this way he produced congenic strains in which the individuals differed from the original parental strain at only a few loci.

Gorer tested Snell's mice with antisera raised against mice that rejected the tumour and found that the congenic strains that rejected the tumour were those which were negative for the antigen II (Gorer et al. 1948). Snell called the genes coding for the tumour resisitance factors 'histocompatibility (H) genes' and the gene coding for the first recognised factor, antigen II, was called 'histocompatibility-2', later abbreviated to H-2, by which the MHC of the mouse is still known.

Two types of H gene were recognised as the result of exchanging skin grafts between Snell's congenic mouse strains. In some cases, the grafts were quickly rejected, while in others, the rejection process was delayed and extended. This indicated the presence in some of the mouse strains of H gene products which had a strong or major effect on tissue rejection and some that had a weak or minor effect. The former became known as the Major Histocompatibility Complex and these classical transplantation antigens became known as the MHC class I antigens (Klein 1979).

The next MHC to be discovered was that of the chicken (Briles et al. 1950), to be followed a few years later by that of the human. The discovery of the human MHC has been attributed to Dausset (1958). Gibson and Medawar (1943) had shown that while skin autografts healed, allografts were rejected and it was assumed that the antigens eliciting the allograft reaction were analogous to those described by Gorer in the mouse. The impetus for the discovery of the MHC in humans was the observations made by Dausset (1954), Payne et al. (1958) and van Rood et al. (1958), of antibodies to leucocyte antigens which caused leukopenia in patients after multiple blood transfusions. A similar phenomemon occurred following blood transfusions of women who, during multiple pregnancies had been immunised with paternal antigens carried on fetal leucocytes which had leaked into the mother's circulation. These leucocyte antigens were presumed to be akin to the mouse H-2 antigens. The human MHC was designated the HLA (human leucocyte antigen) (Dausset, 1958).

The desire to develop a model for allograft reaction *in vitro* and the observed variation in the suscepibility of individuals within an outbred population to infectious disease led to the next major advance, the discovery of the MHC class ll antigens.

To try to simulate the allograft reaction, Bach and Hirschorn (1964) mixed lymphocytes from genetically dissimilar individuals. Although no cell killing was observed, the cells underwent a profound cycle of blast transformation and proliferation which was termed the mixed lymphocyte response (MLR). Subsequently it was shown that the MLR mapped to one end of the H-2 complex which appeared to code for gene products which were not detectable by the available serology (Bach et al. 1972).

In an attempt to dissect the individual variation in the immune response to the complex antigens in pathogens which had been described by various groups (Fink and Quin 1953, Sang and Sobey 1954), Maurer injected rabbits with antigenically simple synthetic peptides and showed that only 40% of animals responded by producing antibody. Subsequently, two inbred lines of guinea pig were identified, one of which responded to the synthetic polypeptide poly-L-lysine, and the other did not (Levine et al. 1963). By crossing these two lines it was shown that the responsiveness was controlled by a single immune response (Ir) gene. Similar studies in mice showed that the Ir gene was linked to H-2 and that the Ir gene, although distinct from H-2, mapped to the middle of the H-2 region (McDevitt and Chinitz 1969, McDevitt et al. 1972, Benacerraf and McDevitt 1972). Hauptfeld et al. (1973) and David et al. (1973) then described a new class of MHC molecules which they considered were the products of the Ir genes and were subsequently called MHC class ll antigens.

The true function of the MHC clearly did not lie in the phenomenology of tumour transplantation and related non-physiological processes. The names linked with the discovery of the physiological function of the MHC molecules are those of Zinkernagel and Doherty. These workers were studying the cytotoxic effects of immune lymphocytes isolated from the spleens of different strains of mice which had

been inoculated intracerebrally with the lymphocytic choriomeningitis virus (LCM). Their *in vitro* test was to look for cell lysis following incubation of the lymphocytes with a virally-infected target cell line. The target cell was the mouse fibroblast L cell which was derived from mice of MHC haplotype H-2^k. They found that only immune lymphocytes from mice which were of the same H-2^k haplotype were able to bring about target cell killing (Zinkernagel and Doherty 1974). This implied some requirement for H-2 compatibility and the authors suggested that the cytotoxic T cells simultaneously recognised viral antigens as being non-self, and the MHC molecules of the target cells as being self. The same workers later identified the specific H-2 component as the MHC class I molecule (Zinkernagel and Doherty 1979).

In this way MHC restriction of antigen presentation to the cells of the immune system was recognised and it will be discussed further following the description of the structure of the MHC, its genes and proteins.

1.2 Genes and proteins of the MHC

The first MHC protein to be sequenced in its entirety was the class I HLA-B7 molecule, and it was sequenced using classical biochemical techniques (Orr et al. 1979). The start of the analysis of the MHC at the DNA level came in 1980 when the sequences of human (Ploegh et al. 1980), and mouse (Kvist et al. 1981) cDNA clones of MHC class I genes were published. Since then the rapid application of the techniques of molecular biology to MHC research has accelerated the identification and sequencing of many gene loci and in some cases many tens of allelic forms at individual loci. In the mapping of the HLA and H-2 complexes, the technique of cosmid cloning proved very useful (Trowsdale et al. 1985, Steinmetz et al. 1986). Relatively large inserts of approximately 40 kbp of genomic DNA could be cloned into cosmid vectors and the genes they contain were linked by producing restriction maps of overlapping cosmid clones (Evans and Wahl 1987). Long range mapping of genes was acheived when the technique of pulsed field gel electrophoresis (PFGE) was developed (Schwartz and Cantor 1984). Conventional agarose gel

electrophoresis permits the resolution of DNA fragments of up to 50 kbp. PFGE on the other hand can resolve fragments up to megabase or even chromosomal size and this has been used in the mapping of the human and the mouse MHC (Hardy et al. 1986, Dunham et al. 1987). A further development has been the introduction of the yeast artificial chromosome (YAC) as a cloning vector (Burke et al. 1987, Schlessinger, 1990). Using these vectors, inserts of hundreds of kbp in size can be cloned . Analysis of the human MHC has now reached the stage that the sequencing of the entire class II region is now imminent.

The arrangement of the numerous genes which map to the MHC in mouse and man has now been established. Schematic representations of the genetic organisation of the HLA and H-2 complexes are shown in Figures 1.1 and 1.2. The HLA map is derived from that of Campbell and Trowsdale (1993) while the H-2 map is derived from Steinmetz et al. (1986). References to the more recently described mouse MHC genes are given in the text below. HLA is contained within a 4000 kbp segment of DNA on the short arm of chromosome 6. The H-2 complex is contained within 2000 kbp on chromosome 17 of the mouse genome. The MHC in both species has been sub-divided into three regions designated class I, II, and III.

1.2.1 The MHC class III region

The group of genes which map between the class II and class I regions in mouse and man (Figures 1.1 and 1.2) form the MHC class III region. This region contains genes which code for components of the complement system and tumour necrosis factor. Klein (1986) considered their position within the MHC as fortuitous as they are not functionally related to components of the class I or II regions. They will not be considered in the present study.

Figure 1.1

A. Map of the human MHC showing the position of the class ll region in relation to class 1 and class lll on chromosome 6.
B. The class ll region expanded to show the positions and orientations of the individual genes. See text for details.

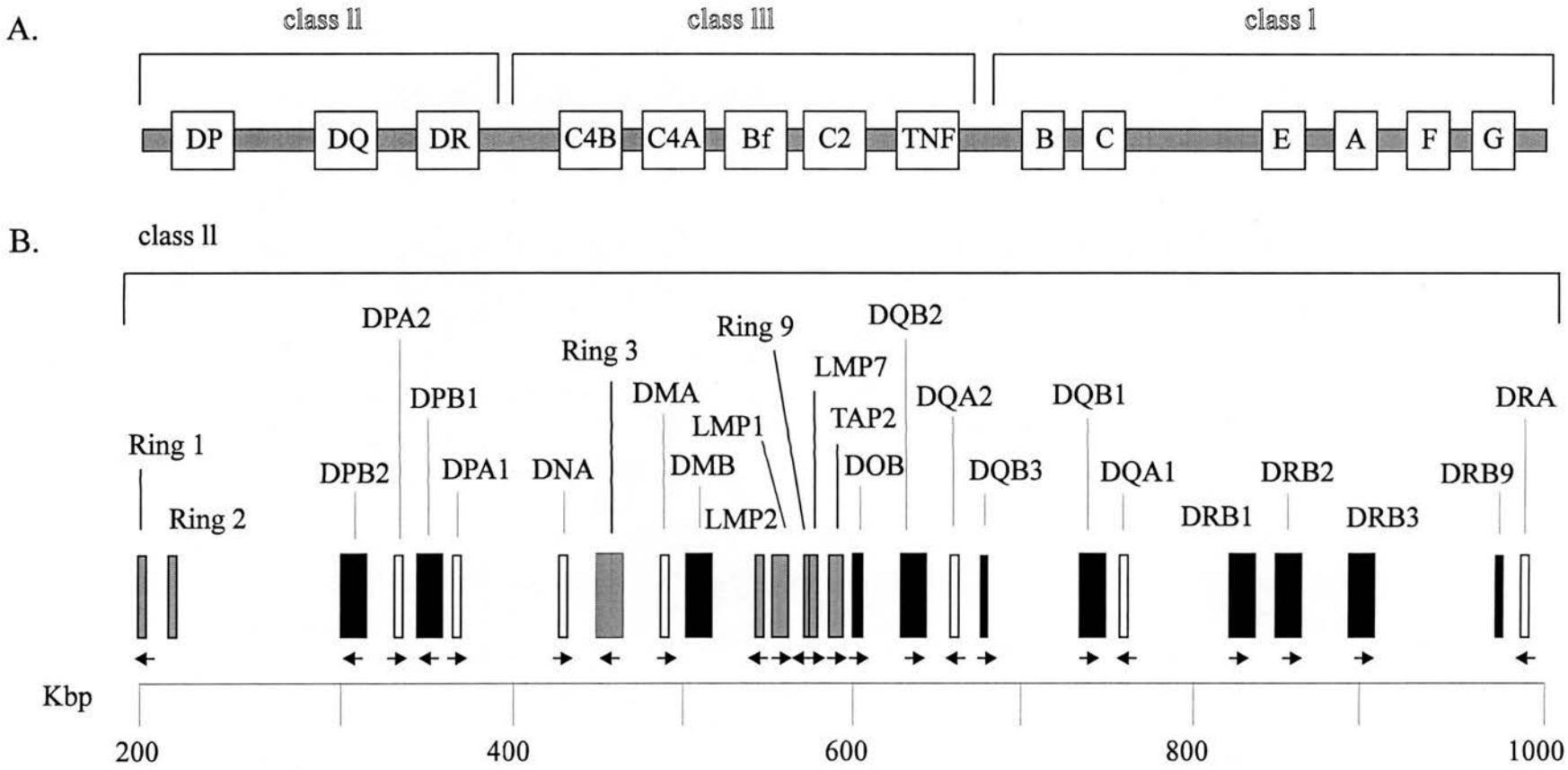
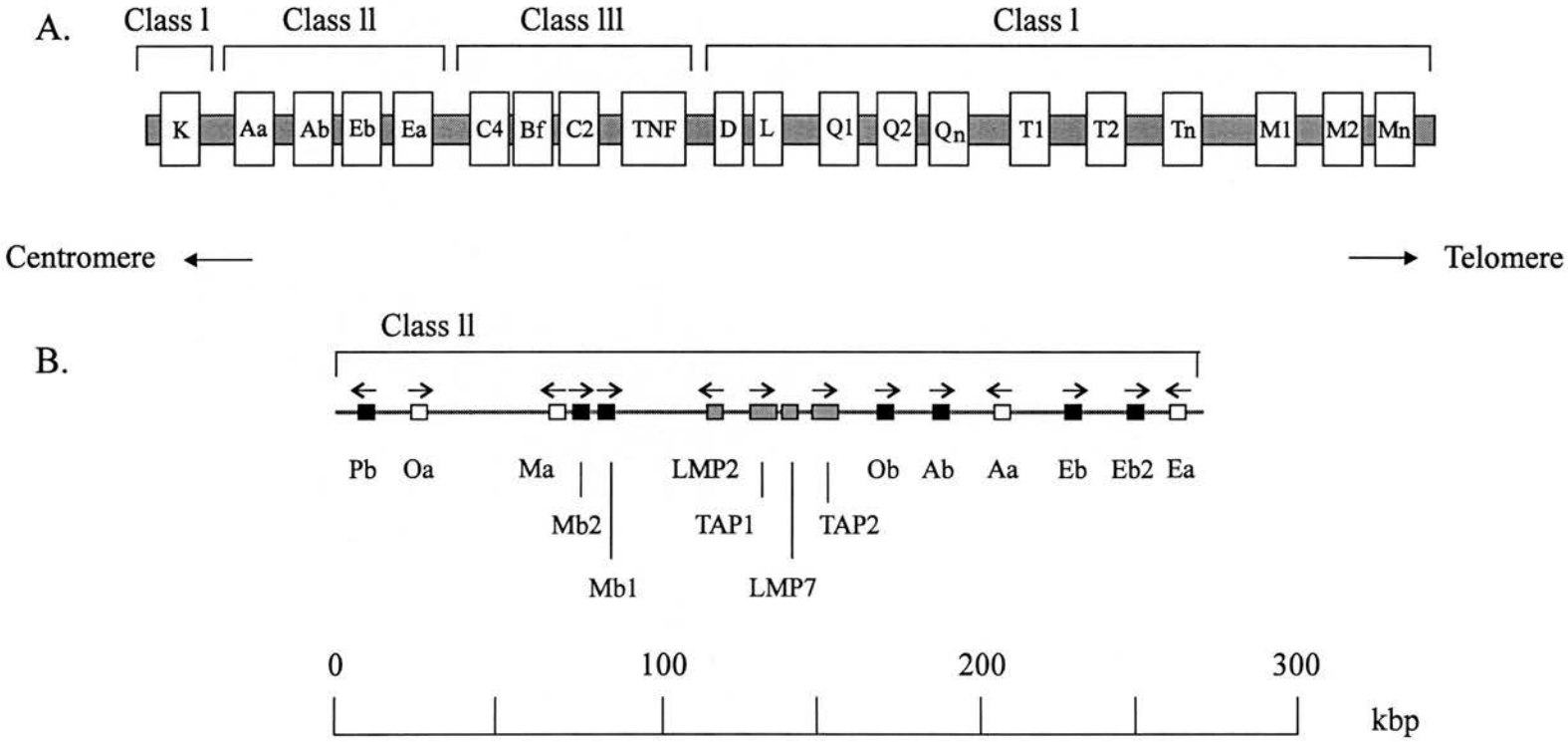


Figure 1.2

Map of the mouse MHC on chromosome 17.

A. The position of the mouse MHC class ll region (H2-I) in relation to the class l and class lll regions.

B. The class ll region expanded to show the postions of the individual genes. See text for details. The scale relates to the expanded class ll region.



1.2.2 The genes and proteins of the class I region

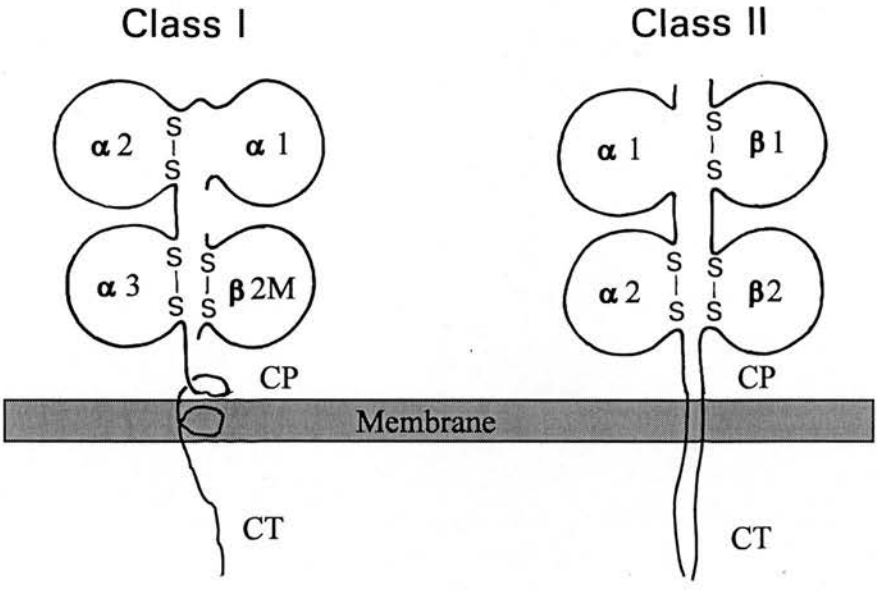
MHC class I genes encode membrane-anchored cell surface glycoproteins of about 44 kDa that are noncovalently associated with β_2-microglobulin (Bjorkman and Parham 1990). The MHC-encoded heavy chain is made up of three extracellular domains designated α1, α2 and α3, a transmembrane domain and a cytoplasmic domain (Figure 1.3). The extracellular domains are made up of about 90 amino acids. The α2 and α3 domains each contain a disulphide bridge which enclose loops of 63 and 86 amino acids respectively. The transmembrane region is made up of about 25 hydrophobic amino acids and serves to anchor the molecule in the membrane. The cytoplasmic region is made of 30-40 hydrophilic amino acids.

The β2-microglobulin is a soluble 12 kDa protein which is not encoded within the MHC. Its gene is located on chromosome 15 in man and on mouse chromosome 2. β2-microglobulin and the membrane-proximal α3 domain share sequence similarities with immunoglobulin domains (Becker and Reeke, 1985, Williams and Barclay 1989). Allelic polymorphism in class I gene products is localised in the α1 and α2 domains, the α3 domain being relatively conserved and the β2-microglobulin invariant (Parham et al. 1988).

There are three loci in both man and mouse which code for the classical transplantation antigens involved in MHC restriction. In the mouse they are termed H-2K, H-2D and H-2L. In man they are termed HLA-A, HLA-B, and HLA-C (Figures 1.1 and 1.2). The rodents appear to be unique in having a class I gene locus at the centromeric end of the complex. All other species appear to follow the HLA model. There are other class I loci in both man and mouse which map telomeric to those of the classical transplantation antigens (Figure 1.2). In the mouse the regions containing non-classical loci have been designated Q, T, and M (reviewed by Shawar et al. 1994). Their sequences are quite dissimilar to those of the classical genes and many of the genes at these loci are probably pseudogenes. However, some of the *H-2T* genes may encode ligands for the γδ T cell receptor (Vidovic et al. 1990), while the *H-2M3*

Figure 1.3

Schematic diagram of the MHC class I and class II molecules. β2M is the β2-microglobulin associated with the class I molecule. The extracellular domains of the different chains of the two molecules are designated α_n and β_n. CP is the connecting peptide. TM the transmembrane domain and CT the intracellular cytoplasmic tail.



gene product specifically presents N-formylated peptides of bacterial origin (Pamer et al. 1992). The biological roles of *HLA-E* and *HLA-F* (Figure 1.1) are uncertain. However, *HLA-G* may have a function at the foeto-maternal interface of the placenta (Kovats et al. 1990).

1.2.3. The genes and proteins of the MHC class II region

The immune response genes (Ir genes) map to the class II region of the MHC. The class II region of the H-2 complex is called the I region while the human class II region is called the HLA-D region. The positions of the class II regions of mouse and man relative to class I and class III are indicated in Figures 1.1 and 1.2, where the class II regions have been scaled up to show the positions of the individual genes.

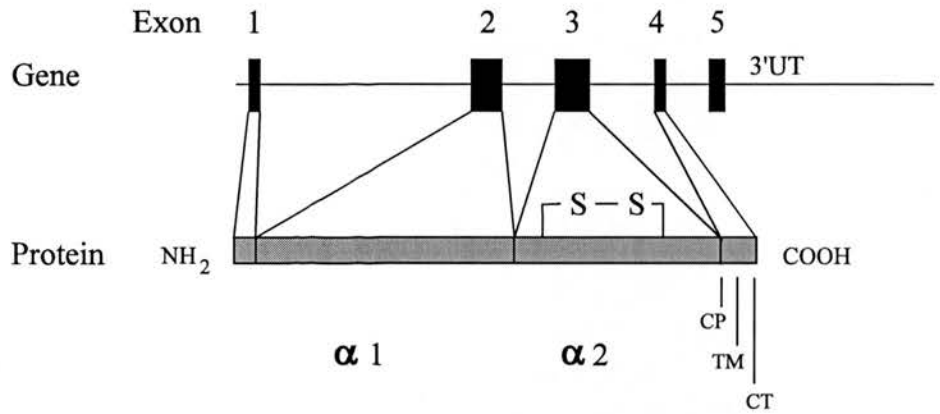
The MHC class II molecule is a type 1 transmembrane glycoprotein. It is a heterodimer made up of a heavy (α), and a light (β) protein chain (Figure 1.3) each encoded by an MHC *A* and *B* gene respectively (Kaufman et al. 1984). The molecular weights of the α chains are in the range 30-34 kDa, while those of the β chains range from 26-29 kDa. Each chain consists of two extracellular domains α1 and α2, and β1 and β2 of approximately 90 amino acids each, a short connecting peptide, a transmembrane region of about 30 amino acids and a cytoplasmic domain (reviewed by Kappes and Strominger 1988). The overall structure of the MHC class II molecule is similar to that of the class I molecule. The class II β1 and β2 domains correspond to the class I α1 and β2-microglobulin domains respectively (Figure 1.3). The class II α2 and β2 domains, like the class I α3 and β2-microglobulin domains show structural similarities to Ig domains (Kaufman 1984). The MHC class I and class II molecules belong to the immunoglobulin superfamily.

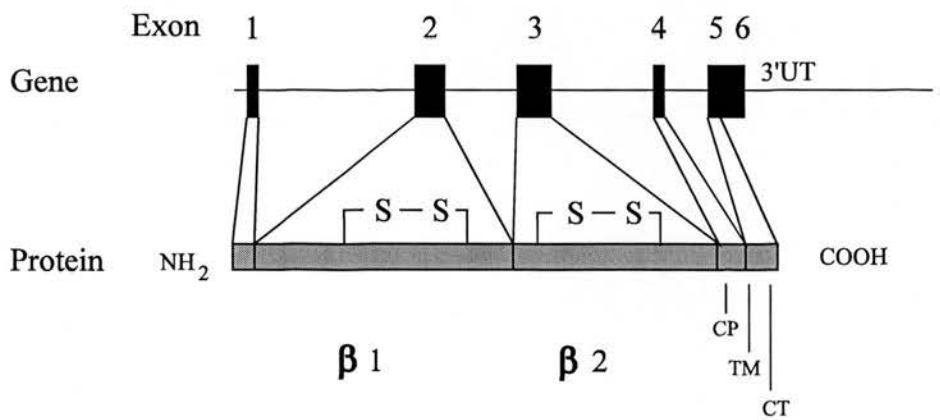
A schematic drawing of the structures of the class II *A* and *B* genes is shown in Figure 1.4. When the class II glycoproteins are translated, they are equipped with an N-terminal signal sequence responsible for the translocation of the molecule into the endoplasmic reticulum (Kappes and Strominger 1988). This sequence is encoded by

Figure 1.4

Schematic showing the structure of the gene and the protein of the MHC class ll molecule. Exons are represented by black boxes and the protein domains by stippled boxes.

L is the leader sequence, CP the connecting peptide, TM the transmembrane region, CT the cytoplasmic tail and 3'UT the untranslated region.





exon 1. Exons 2 and 3 encode the membrane-distal and -proximal domains respectively. Exon 4 encodes the transmembrane and cytoplasmic tail of α chains, but only part of the cytoplasmic domain in β chains, the remainder being encoded by exon 5. Exon 5 in *A* genes contains the 3' un-translated region which is contained in exon 6 of *B* genes.

In man there are now five recognised class II sub-regions. The *DP, DN/DO, DQ* and *DR* (reviewed by Kappes and Strominger 1988) and the more recently described *DM* sub-region (Kelly et al. 1991). The *DP* and *DQ* sub-regions each contain an expressed *A/B* gene pair while the *DR* sub-region contains an *A* gene which is expressed together with one or two *B* genes depending upon the haplotype (Rollini et al. 1985, Andersson et al. 1987). The *DP* sub-region also contains two pseudogenes *DPA2* and *DPB2* (Gustafsson et al. 1987). The *DQ* sub-region contains a gene fragment, *DQB3* (Ando et al. 1989), and another *A/B* gene pair, *DQA2* and *DQB2,* which although they are apparently not pseudogenes, are not expressed (Jonsson et al. 1987). The *DN/DO* sub-region is represented by single, well separated, non-polymorphic *A* and *B* genes, *HLA-DNA* and *-DOB* (Trowsdale and Kelly, 1985, Young and Trowsdale 1990, Tonnelle et al. 1985, Servenius et al. 1987). Although these genes are transcribed, no protein product or function has been assigned to them. The *DMA* and *DMB* loci map between the *DNA* and *DOB* genes (Kelly et al. 1991a). Although they are class II-like, their sequences are quite divergent. Recent information has shown that the *DM* protein has an accessory role in peptide loading of class II molecules (Sloan et al. 1995). Other genes which map to this area of the HLA-D region are not related to MHC class II genes but encode proteins which are intimately involved in antigen presentation to $CD8^+$ cytotoxic T cells which is mediated by MHC class I molecules. These are the LMP (Low molecular mass polypeptide) and TAP (Transporters associated with Antigen Processing) genes indicated in Figure 1 (reviewed by Monaco 1992).

The mouse class II or *H-2I* region contains orthologues of many of the genes of the *HLA-D* region (Steinmetz 1986). The mouse genes corresponding to the expressed

human *DQ* and *DR* genes are *Aa/Ab* (Kaufman et al. 1984) and *Ea/Eb* (Denaro et al. 1985) respectively. The *H-2E* sub-region also contains a gene, *Eb2*, which is transcribed but it is not known whether or not it is translated (Denaro et al. 1985). The mouse gene *Pb* (formerly called Aβ3), corresponds to an *HLA-DPB* gene but is a pseudogene (Widera and Flavell 1985). The mouse *Oa* and *Ob* (formerly Aβ2) genes are interesting in as much as they correspond to the *HLA-DNA* and *HLA-DOB* genes and have been shown to be expressed but with a restricted tissue distribution (Larhammar et al. 1985, Karlsson et al. 1991, 1992). The region between *Oa* and *Ob* contains genes which correspond to the DM gene loci, Ma, Mb1 and Mb2 (Cho et al. 1991a). The same region contain orthologues of the genes involved in class I antigen presentation LMP and TAP (Cho et al. 1991b).

1.3 Expression of MHC molecules

MHC class I is expressed at high levels on the surface of cells of the immune system and on most somatic cells (Klein 1975). However, class I is not found on certain cells of the pancreas, neurons of the CNS, or cells of the villous trophoblast (Daar et al. 1984a, b). MHC class II molecules on the otherhand are more restricted in their cellular distribution. They are primarily expressed on cells involved in the presentation of antigen to the $CD4^+$ subset of T helper lymphocytes (T_h), such as monocytes, macrophages (Unanue, 1984), dendritic cells (Brooks and Moore, 1988), and B cells (Kearney et al. 1977). T cells can express class II following activation by alloantigen or mitogen (Metzgar et al. 1979). Interferon gamma (IFN-γ) increases the expression of MHC class II on some of these cell types (Houghton et al. 1984, Collins et al. 1986), while interleukin 4 (IL4) increases class II expression on resting B cells (Polla et al. 1986). The regulation of class II expression is complex (Glimcher and Kara, 1992). Not all class II loci are co-ordinately expressed. For example, *HLA-DOB, H-Ob* and *H-2Eb* are not inducible by IFN-γ (Tonnelle et al. 1985, Wake and Flavell 1985).

1.4 Function of the MHC

The primary function of the MHC molecule is to present pathogen-derived peptide antigens to T cells (Germain and Margulies 1993). MHC class I molecules present processed antigen to the cytotoxic T cell subset (T_c) which express the CD8 co-receptor , a secondary, peptide independent, class I-binding molecule. CD8 interacts with conserved peptide motifs in the α3 domain of the MHC class I (Salter et al. 1990). This interaction is profoundly affected by amino acid changes in the α3 domain of class I (Potter et al. 1989). MHC class II presents processed antigen to the $CD4^+$ T_h subset. CD4 interacts with the membrane-proximal β2 domain of the class II molecule (Konig et al. 1992) which is homologous to the α3 domain of class I.

Class I and class II molecules present peptides derived from different sources. Class I molecules present peptides derived from endogenous or cytosolic proteins (Townsend et al. 1985, 1986, 1989, Jardetsky and Lane 1991). These include self proteins, as well as proteins derived from viral mRNAs or from bacteria or protozoa which have penetrated into the cytosol. The signal resulting from the recognition of a class I/peptide complex by the $CD8^+$ T cell receptor induces a cytolytic response, thus killing infected cells.

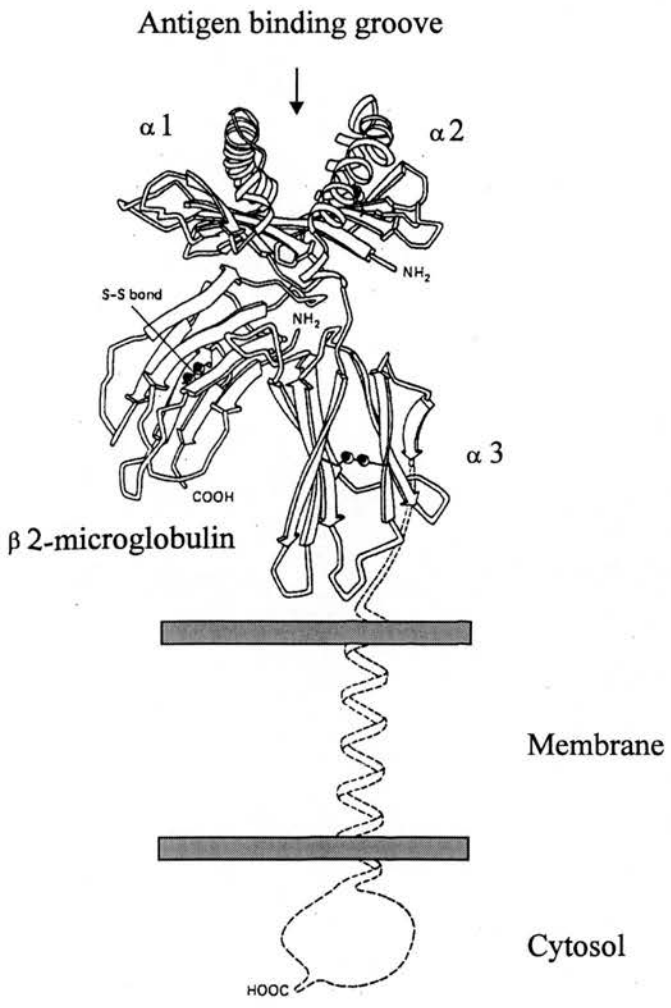
Class II molecules present peptide derived from exogenous proteins. These include soluble proteins or those associated with whole organisms which have been phagocytosed. The signal resulting from the recognition of a class II/peptide complex by the CD4+ T helper cell antigen receptor is a cytokine cascade (IL2, 4, 5 and IFN-γ) which induces and regulates T and B cell proliferation and function, macrophage activation and B cell differentiation. The latter develop into plasma cells which generate antigen-specific antibodies. Recent studies have revealed much of the intricate and different routes taken by the MHC molecules within cells to receive and present their peptide loads.

The class I and class II genes are highly polymorphic. However, in any one individual, assuming heterozygosity at all of the loci and allowing for the fact that the MHC genes are co-dominantly expressed, only a small number of MHC molecules are available to present the broad spectrum of peptide antigens for T cell recognition. Each molecule must therefore present a large array of peptides. This necessitates a compromise between high affinity and broad specificity of peptide binding. The binding affinity between the MHC molecule and the peptide must be high enough to allow the MHC molecule to retain the peptide at the cell surface long enough to allow recognition by the T cell antigen receptor in the face of an effective unbound peptide concentration of zero.

The relationship between structure and function was brought into sharp focus by the elucidation of the three dimensional structure of the MHC class I molecule by X-ray crystallography (Bjorkman et al. 1987) and a schematic of the MHC class I molecule is shown in Figure 1.5. This also allowed the prediction of the tertiary structure of the class II molecule (Brown et al. 1988) which was subsequently shown to be accurate by direct determination of the class II crystal structure (Brown et al. 1993). The structures show that in the case of MHC class I, the $\alpha 1$ and $\alpha 2$ membrane-distal domains of the heavy chain each contribute four strands to an eight stranded anti-parallel β sheet. They also contribute one of two interrupted anti-parallel α-helices which overlie and push against the side of the β sheet. The α helices are separated by an extended groove 30 angstroms long, which is 12 angstroms wide in the middle, tapering to 5 angstroms at each end. The $\alpha 1$ and $\alpha 2$ domains of the class I molecule are highly polymorphic (Lawlor, 1990). The highly variable amino acid residues point into this peptide binding groove. Highly conserved bulky amino acids tyrosine and tryptophan at positions 84 and 167 block each end of the groove (Bjorkman and Parham 1990). Potentially antigenic peptides of 8-9 amino acids in length (Madden et al. 1992) are tightly bound within the groove by the interaction between the side chain groups of the MHC amino acid residues along the groove and the backbone of the bound peptide. In addition the side chains of the peptide amino acids are bound within side pockets formed in the groove. The N- and C-termini of the peptides are buried

Figure 1.5

A schematic diagram of the structure of the MHC class 1 molecule. The diagram has been modified from that of Alberts et al. (1989) which was based on the X-ray crystallographic data of Bjorkman et al. (1987).



and interact with highly conserved MHC residues at the end of the groove. This has the effect of fixing the N-C orientation of the peptide in all class I/peptide complexes. The extreme polymorphism of the class I genes is therefore reflected in the different peptide-accomodating specifications of the antigen-binding groove.

The peptide binding groove of the class II molecule is formed from an interchain dimer of the $\alpha 1$ and the $\beta 1$ domains. The α-helix contributed by the $\alpha 1$ domain of the class II molecule is shorter than the corresponding α-helix from the class I $\alpha 1$ domain and sits closer to the floor of the groove. The conserved amino acids at the ends of the groove are smaller than in the class I structure, arginine versus tyrosine and asparagine versus tryptophan. This has the effect of creating an open-ended, wider groove which can accomodate peptides up to 25 amino acids in length (Brown et al. 1993).

The biological specificity of peptide loading of MHC class I and class II molecules is brought about in different intracellular compartments. The processing of peptide destined for class I mediated presentation involves the proteasome, a large, 700 kDa protease of multiple subunits, two of which LMP2 and LMP7 , are encoded in the class II region (Figures 1 and 2) (Goldberg and Roche 1992). The peptides are translocated to the endoplasmic reticulum (ER) by specific transporters encoded by the class II region TAP genes (Monaco 1992). Newly synthesised class I and $\beta 2$-microglobulin molecules enter the ER where class I associates with a chaperone molecule, calnexin. Peptide and $\beta 2$-microglobulin bind, the complex is released from calnexin, transported through the Golgi apparatus and to the cell surface (Neefjes et al. 1990).

The α and β subunits of the class II molecules have an N-terminal signal sequence responsible for their translocation into the ER (Kappes and Strominger 1988). The signal sequences are cleaved and the chains are integrated into the ER membrane. The class II chains then become intimately associated with another molecule, the invariant (Ii) chain (Jones et al. 1979). Ii is a non-polymorphic type II transmembrane glycoprotein (N-terminus in the cytoplasm, COOH terminus in the lumen of the ER).

It is 216 amino acid long and occurs as trimers in the absence of class II (Marks et al. 1990). Following Ii trimerisation, three class II αβ dimers bind to form an αβIi nonamer. This complex is then translocated through the Golgi stacks into the trans-Golgi where they then deviate from the exocytic pathway that takes most proteins including class I to the cell surface. Instead they move to the endocytic pathway (Neefjes et al. 1990). The proteolytic destruction of Ii is a pre-requisite for peptide loading of class II. The class II/peptide complex is then transported to the cell surface. Recent results have extended our knowledge of this process to show that during proteolysis, a peptide derived from Ii (CLIP, class II associated invariant chain peptide) remains tightly bound to the class II binding site (Riberdy et al. 1992) and it appears that the products of the *DM* class II locus (Figures 1 and 2) bring about the release of CLIP thus freeing the binding site for antigenic peptides (Sloan et al. 1995).

1.5 The sheep MHC

At the start of the project described here therefore, there was a wealth of information on the detailed structure and function of the MHC of mouse and man. Much less was known about other species (reviewed in Klein 1986). Restriction fragment length polymorphism (RFLP) studies in cattle had demonstrated the presence of orthologues of *HLA-DQ, DR, DN* and *DO* genes but not *DP* (Andersson 1988, Andersson and Rask 1988, Andersson et al. 1986a,b, 1988, Sigurdardottir et al. 1988). Furthermore, the presence in cattle of non-orthologous genes designated *DY* which segregated with *DOB* in a group separated by a recombination frequency of 0.17 from the region which contains the *DQ* and *DR* genes had been noted (Andersson et al. 1988). However, the only nucleotide sequences of cattle class II genes reported were of two *DRB* pseudogenes (Muggli-Cockett and Stone, 1988, 1989).

Alloantisera had been raised to sheep class I molecules by immunising animals with allogenic lymphocytes followed by analysis using a cytotoxicity assay (Millot, 1974, 1979, Ford 1974, 1975). Serology indicated the presence of three class I loci containing up to 16 alleles (Millot, 1979, 1984). These were designated OLA (Ovine

Leucocyte Antigen) -A, B and C. (Thoughout this thesis, the sheep MHC will be referred to as *Ovar*, from the first two letters of the specific name *Ovis aries* following the nomenclature system of Klein et al. 1990). Immunoprecipitation of sheep class I molecules from lysates of lymphocytes using anti-sheep class I monoclonal antibodies, demonstrated that like the human and mouse proteins, the sheep molecules were heterodimers consisting of a heavy alpha chain of 44 kDa and a 12 kDa β2-microglobulin (Gogolin-Ewens et al. 1985, Puri et al. 1987a). The tissue distribution of sheep class I molecules was similar to that of mouse and man.

Two groups have produced monoclonal antibodies to sheep class II molecules (Hopkins et al. 1986, Dutia, 1990; and Puri et al. 1985, 1987 a,b,c, Puri and Brandon 1987). Immunoprecipitation and SDS-PAGE gels showed that the sheep class II molecule was a heterodimer consisting of a heavy chain of 32-36 kDa and a light chain of 24-29 kDa. The tissue and cellular distribution of sheep class II molecules was similar to that of mouse and man ie monocytes, macrophages, B cells, dendritic cells. As in the human, sheep class II was also expressed on activated T cells.

At the DNA level, Chardon and others (1985) had demonstrated by Southern blotting that HLA class II gene probes hybridised to sheep DNA. On the basis of differential hybridisation to HLA gene probes, *Ovar-DQA* and *DRA*-like genes, together with a number of undefined sheep class II *B*-like genes were obtained from a genomic bacteriophage λ library (Scott et al. 1987).

Clearly, the polymorphic amino acids of the MHC molecule are crucial in determining whether or not a specific peptide will be presented T cells. Similarly, they are important in the selection of T cell receptor specificities during the process of T cell differentiation and maturation in the thymus. It is therefore not surprising that many diseases involving immunological dysfunction in man, the autoimmune diseases such as insulin dependent diabetes mellitus (IDDM), rheumatoid arthritis, ankylosing spondylitis (AS) and coeliac disease are associated with certain MHC alleles (Batchelor and McMichael 1987, Nepom and Erlich 1991). The strongest association

occurs between AS and the HLA-B27 molecule which is present in approximately 90% of patients. However, as greater than 95% of the HLA-B27+ population do not develop the disease (Derhaag, 1990), there are clearly a number of genetic and environmental factors involved. A form of molecular mimicry between HLA-B27 and motifs from the arthritogenic bacteria has been proposed but was not supported by experiment (Kapasi et al. 1992, Lahesmaa et al. 1992). In the case of IDDM the strong association with HLA-DR3 and DR4 (Horn et al. 1988) has been localised to alleles which code for a non-aspartic acid residue at position 57 of *DQB1* and with *DQA1* alleles which have an arginine at position 52 of the protein chain (Tosi et al. 1994).

Associations have also been sought between MHC and suscepibility or resistance to infectious disease, but rarely identified. Susceptibility to typhoid and yellow fever (De Vries et al. 1989) and malaria (Hill et al. 1991, 1994) has been shown to be HLA-associated. In farmed animals, the enormous economic loss due to infectious disease and its control, has prompted the search for genetic markers associated with disease resistance for use in marker-assisted host selection programs. Associations between MHC class I and diseases ranging from mastitis to nematode infection have been reported in cattle (reviewed by Ostergard et al. 1989). In sheep, associations between the genetic variation in resistance to nematode infection and the MHC have been sought. However, the results have been equivocal. While Cooper et al. (1989) found no effect of MHC on resistance to *Haemonchus contortus,* an MHC class I influence on the immune response to *Trichostrongylus colubriformis* has been suggested (Outteridge et al. 1985, 1988), although this was not consistent (Outteridge et al. 1986).

1.6 Aims of the present study

Despite the wealth of information available about the MHC of mouse and man, the study of the MHC of an economically important group of animals such as the ungulates at a fundamental level can be justified for a number of reasons. Without

doubt a major impetus has been the desire to improve the disease resistance of herds and flocks alluded to above. The availability of species-specific and locus-specific probes and reagents should facilitate such studies. Because of its highly polymorphic, multi-locus nature, there has been much interest in the mechanisms by which the MHC has evolved. Access to the genetic structure and gene sequences of a major and disparate group such as the ungulates who last shared a common ancestor with mouse and man approximately 70 million years ago, is clearly of interest to such studies. Furthermore, the sheep because of its size and temperament, is a convenient species in which to study aspects of fundamental immunology using surgical intervention techniques such as cannulation of lymphatic ducts to study the dynamics of immune responses *in vivo*. Once again such studies are facilitated by species-specific and locus-specific probes and reagents from the immunologically important MHC. These will allow an analysis of the spectrum of MHC restriction in presenting pathogen peptides to T cells.

The objectives of the present study were therefore to map the class II region of the sheep MHC, to identity the similarities and differences between the MHC of the sheep and the well-defined MHCs of mouse and man, and to identify expressed genes with the longer term aim of undertaking functional studies.

The approach taken for the study of the sheep class II region described here was the characterisation of cosmid clones obtained from genomic libraries constructed using cosmid vectors and screened with class II gene probes from mouse and man. Cosmid vectors were used for two main reasons. They facilitate genomic mapping because they permit the cloning of relatively large, (up to 45 kbp), pieces of DNA. Secondly, they can be used in transfection studies to identity sheep class II *A* and *B* genes which express the corresponding class II molecule at the cell surface. The cell line used for this purpose was the mouse L cell, a class II negative fibroblast. The reactivity of two panels of anti-sheep class II monoclonal antibodies with the expressed class II molecules on the L cells could then be used to further define their isotype specificities. Class II genes are relatively large, the *B* genes are approximately 9 kbp in length, and

so the probability of finding clones containing complete genes and expressible *A/B* gene pairs was greatly enhanced by the use of cosmids.

Two genomic libraries had been constructed in the Department of Immunology at the Babraham Institute, Cambridge, by E.V. Deverson, N. Huskisson and S.Watson. The libraries were screened and a large number of clones isolated. I became involved in the screening of these clones with human and mouse MHC class ll gene probes and their subsequent restriction mapping. Those clones identified by a number which includes a decimal point eg 46.1, were mapped by E. Deverson. Thirty one clones were isolated which apparently contained seven distinct *A* genes, and 24 different *B* genes or gene fragments Deverson et al. 1991). A major part of this thesis describes the further characterisation of these genes. In a parallel study, Ballingall et al. (1992) co-transfected all combinations of the *A* and *B* genes into mouse L cells to find expressed pairs. The only pair which formed a cell surface molecule came from two different cosmids. The *A* gene appeared to be *DRA*-like on the basis of its stronger hybridisation to an *HLA-DRA* probe. The *B* gene was sequenced and proved to be of the *DR* isotype. As it was very unlikely that *DQ* genes were not expressed in the sheep, a third cosmid library was constructed and the subsequent screening for expressed *DQ*-like genes is described in Chapter 3. The screening of the third library also produced clones containing MHC class ll genes which had not been described previously in the sheep. The last two sections of the present study describe the characterisation of these genes.

Chapter 2

Materials and methods

2.1 Bacteriological techniques

2.1.1 *E. coli* strains

The following *E. coli* strains were used in this study.

Strain	Genotype	Reference
JM109	*rec*A1 *sup*E44 *end*A1 *hsd*R17 *gyr*A96 *rel*A1 *thi*-D(*lac*-proAB) F' [*tra*D36 proAB$^+$ *lac*Iq *lac*ZΔM15]	Yannisch-Perron et al 1985
DH1	*rec*A1 *sup*E44 *end*A1 *hsd*R17 *gyr*A96 *rel*A1 *thi*-1	Hanahan 1984
490A	*rec*A1	

2.1.2 Growth media

All quantities are based on 1 litre of solution.

L-broth	10g tryptone, 5g yeast extract, 5g NaCl
L-agar	L-broth containing 15g agar l^{-1}
Top agar	L-broth containing 7g agar l^{-1}
2xTY broth	16g tryptone, 10g yeast extract, 5g NaCl
2xTY agar	2xTY broth containing 15g agar l^{-1}
M9 minimal agar	40mM Na_2HPO_4, 20mM KH_2PO_4, 20mM NH_4Cl, 9mM NaCl, 1mM $MgSO_4$, 0.1mM $CaCl_2$, 1mM thiamine-HCl, 2g glucose, 15g agar.

The following antibiotics were added to growth media as required and at the stated concentrations.

Antibiotic :	Tetracycline	Ampicillin	Kanamycin
Concentration μg ml^{-1} :	12.5	50	50

2.1.3 Preparation of competent *E. coli*

A working stock of *E. coli* JM109 was maintained at 4°C on a minimal agar plate to ensure retention of the F episome required for the detection of recombinants by colour discrimination (section 2.2.18.2). 10 ml of L broth were inoculated with a single colony shaken overnight at 37°C. 1 ml of this culture was added to 500 ml of L broth and the bacteria grown to an optical density of 0.9 at 600 nm. The cells were pelleted at 4000xg for 10 minutes and resuspended in 250ml of sterile ice cold 100mM $CaCl_2$. The cells were left on ice for 30 mins and re-centrifuged. The pellet was resuspended in 5ml of ice-cold 100mM $CaCl_2$ containing 15% glycerol. Aliquots were snap frozen and stored at -70°C.

2.2 Molecular biology techniques

2.2.1 Cloning vectors

The following cloning vectors were used in this study.

Vector	Type	Size (kbp)	Markers	Reference
pBS+	Phagemid	3.2	amp *lacZ'*	Yannich-Perron et al. (1985)
pBScript SK(+)	Phagemid	2.9	amp *lacZ'*	Short et al. (1988)
pTL5	cosmid	5.7	tetracycline	Steinmetz et al. (1982)
pTL6	cosmid	5.2	tetracycline	Steinmetz et al. (1982)
pCos8	cosmid	6.7	kanamycine	Ehrlich et al. (1987)
M13mp18/19	phage	7.2	*lacZ'*	Messing (1983)

2.2.2 Preparation of cosmid and plasmid DNA.

Large amounts of cosmid and plasmid DNA, 500-1000 µg, were prepared by the alkaline extraction technique described by Birnboim (1979), followed by purification

on caesium chloride density gradients. 10ml of 2xTY broth containing the appropriate antibiotic were inoculated incubated with vigorous shaking at 37°C for a few hours. The cultures were added to 2 litre indented conical flasks containing 250ml of the same medium and incubated overnight with vigorous agitation. The cells were collected by centrifugation at 4000xg for 10 minutes and were resuspended in 100ml ice cold glucose buffer (50mM glucose, 25mM Tris/HCl pH8, 10mM EDTA). After centrifugation at 4000xg for 10 minutes, the cells were resuspended in 20ml of glucose buffer, lysed by the addition of 40 ml of freshly prepared 0.2 M NaOH, 1% w/v SDS and left for 5 minutes on ice. Bacterial chromosomal DNA and contaminating proteins were removed by precipitation following the addition of 30 ml of 3 M potassium acetate, 1.8 M formic acid. The mixture was held on ice for 30 minutes and the precipitate removed by centrifugation for 10mins at 15000xg. The nucleic acids were precipitated by the addition of 0.6 volumes of isopropanol to the supernatant and pelleted by centrifugation at 15000xg for 15 minutes. The pellet was washed in 70% ethanol and dissolved in 9 ml of TE. Nine grams of caesium chloride were added, followed by 1ml of ethidium bromide solution (10mg ml^{-1}, Sigma Chemical Co, Cat. no. E8751). The solution was transferred to a 10ml "Quick Seal" ultracentrifuge tube (Beckman Instruments Ltd, High Wycombe, England, Cat No 342413) and centrifuged in a 70Ti rotor on a Beckman L8M ultracentrifuge for 36 hours at 40000 rpm. The band of cosmid DNA was clearly visible in daylight and was removed from the tube using a 1ml syringe fitted with an 18G needle. The ethidium bromide was removed by extraction into water-saturated *n*-butanol. The DNA solution was diluted 1:1 with water and the DNA precipitated by the addition of two volumes of ethanol. The DNA was pelleted, washed with 70% ethanol and dissolved in TE at a concentration of 1μg μl^{-1} determined spectrophotometrically

2.2.3 Small scale preparation of cosmid and plasmid DNA (mini-prep)

This method was used for the rapid isolation of small amounts, 10-15 μg of plasmid or cosmid DNA. The technique was as described in section 2.2.2 except on a reduced scale, and omitted the 36 hour ultracentrifugation stage. A 10ml culture was

incubated overnight at 37°C. Following centrifugation for 10 minutes at 4000xg, the supernatant was discarded, the cells were resuspended in 1ml of ice cold glucose buffer and transferred to a 1.5 ml microcentrifuge tube. The cells were pelleted at 13000 rpm for 10 seconds and resuspended in 200 µl cold glucose buffer. 400 µl of SDS/NaOH solution were added and cells left on ice for 5 minutes. This was followed by the addition of 300 µl of potassium acetate/formic acid solution and the mixture was left on ice for a further 30 minutes. After centrifugation for 5 minutes at 13000 rpm, the supernatant was transferred to a fresh tube and contaminating protein removed by extraction with an equal volume of phenol/chloroform (1:1). The tube was centrifuged at 13000 rpm for 5 minutes. The top, aqueous phase was removed to a fresh tube and re-extracted with chloroform. Following re-centrifugation, the aqueous phase was again removed to a new tube and the DNA precipitated by the addition of 0.6 volumes of isopropanol. The DNA was pelleted by centrifugation, washed in 70% ethanol, and dissolved in 50 µl TE.

2.2.4 Quantification of DNA and RNA

Nucleic acid was quantified by spectrophotometry. DNA and RNA have characteristic UV spectra, the peak of absorbance is at 260 nm, while the absorbance at 280mn is a measure of the amount of contaminating protein which is present (Sambrook et al 1989). The absorbance at 260nm gives a measure of the concentration of the nucleic acid sample. An absorbance of 1 ODU_{260} corresponds to a concentration of 50 µg ml^{-1} for double stranded DNA and 40 µg ml^{-1} for RNA. The ratio OD_{260}/OD_{280} is approximately 1.8 and 2.0 for clean DNA and RNA samples respectively.

2.2.5 Restriction endonuclease digestion of DNA

The restriction enzymes were purchased from Boehringer Mannheim and used with the accompanying buffers following the manufacturer's instructions. The final concentration of enzyme was usually 10 units μg^{-1} of digested DNA.

2.2.6 Ligation of DNA

The enzyme T4 DNA ligase catalyses the formation of a phosphodiester bond between a 5' phosphate group and a 3' hydroxyl group resulting in the covalent linking of two DNA molecules which have compatible ends produced by restriction enzyme digestion. Ligation reactions were set up in 20 µl total volume containing 100 ng of vector, insert DNA at a ratio of insert : vector of 4 : 1 where possible, 2 µl of 10x ligase buffer (660mM Tris/HCl, 50mM $MgCl_2$, 10mM DTT, 10mM ATP, pH 7.5) and 2 u of T4 ligase (Boehringer Mannheim). Reactions were incubated at 37^0C for 2 hours or at 12^0C overnight.

2.2.7 Dephosphorylation of DNA

Calf intestinal alkaline phosphatase (CIP), hydrolyses the terminal 5' phosphate groups from DNA molecules, so inhibiting self-ligation. DNA samples were dissolved in 84 µl of water, 10 µl of 10x CIP buffer (0.5M Tris/HCl pH 9.0, 10mM $MgCl_2$, 1mM $ZnCl_2$, 10 mM spermidine), and 6 units of CIP were added. The reaction was incubated at 37^0C for 30 minutes when a further 6 units were added and the incubation repeated. The sample was phenol extracted and ethanol precipitated before further processing.

2.2.8 Transformation of competent *E. coli*

A 50 µl aliquot of the frozen stock of competent bacterial cells (section 2.2.3) was thawed on ice. 5 µl of the ligation reaction (approximately 50 ng of DNA) was added

and the mixture held on ice for 30 minutes, after which it was heat-shocked at 42^0C for 2 minutes. The cells were spread on L-agar plates containing ampicillin, and X-gal and IPTG for colour selection of recombinants (see section 2.2.18.2). The plates were incubated at 37^0C overnight. Colourless recombinant colonies were inoculated into 10 ml of L broth and DNA prepared as in sections 2.2.1 or 2.2.2.

2.2.9 Agarose gel electrophoresis of cosmid or plasmid DNA

DNA fragments resulting from restriction digests of either cloned DNA or genomic DNA were analysed by conventional agarose gel electrophoresis. Agarose concentrations ranged from 0.3-1.0 % w/v depending on the size of the DNA to be resolved. Gel dimensions were either 10x7 cms or 20x20 cms. Electrophoresis buffer was either TBE (45mM Tris/borate, 1mM EDTA pH7.3) or TAE (40mM Tris/acetate, 1mM EDTA pH7.2) and gels were run at 25 volts overnight or 75 volts for one hour. Ethidium bromide was added to a final concentration of 0.5 $\mu g\ ml^{-1}$ and the DNA fragments were visualised by their fluorescence under UV illumination. A permanent record was obtained by photographing the gels on Polaroid 667 film (Polaroid UK Ltd, St Albans, England) or on heat sensitive paper (Mitsubishi video copy processor, UVP). The size of the DNA fragments was determined by comparison with known standards generated by digestion of λ phage DNA with HindIII or other suitable restriction enzymes.

2.2.10 Purification of DNA fragments from agarose gels

To prepare a fragment of DNA for use as a probe in hybridisation experiments or for subsequent subcloning, the fragment was separated on an agarose gel and then recovered from the gel by adsorption onto powdered silica under conditions of high salt concentration as described by Vogelstein and Gillespie (1979). A commercially available kit was used (Geneclene, Stratech Scientific Ltd, Luton). The fragment of agarose containing the DNA of interest was excised from the gel and melted in three volumes of the chaotropic agent 6M NaI at 55°C. The silica matrix was then added to

bind the DNA. After 5 minutes on ice, the silica was pelleted at 13000 rpm and washed three times with an alcohol solution held at -20°C. The DNA was recovered from the matrix by resuspending the pellet in 20 µl of water, incubation at 55°C for 2 minutes, followed by centrifugation to remove the silica.

2.2.11 Analysis of DNA by Southern blotting

After electrophoresis in agarose gels, both cloned and genomic DNA was transferred to Hybond-N membranes (Amersham International, UK), using a modification of the method of Smith and Summers (1980).

The DNA was depurinated by soaking in 0.25M HCl for 10 minutes followed by denaturation in 0.5 M NaOH/1.5 M NaCl for 30 minutes. The gel was transferred to a 1M ammonium acetate/0.02M NaOH solution. After 1 hour, the gel was removed and a piece of Hybond-N, cut slightly bigger than the gel, was placed on top of the gel. Three pieces of 3MM filter paper were cut to the same size as the gel, soaked in the ammonium acetate and placed on top of the membrane. A 10 cm thickness of paper towel was positioned on top of the filter paper and finally a weight was placed on top of the towels. The blot was left overnight for the DNA transfer to take place by capillary action. The next day, the membrane was air dried and the DNA UV-fixed.

2.2.12 Analysis of RNA by Northern blotting

A slightly different procedure was followed for the electophoretic separation of RNA and its transfer to nylon membranes. Total RNA was prepared as described in section 2.3.1. A denaturing 1% agarose gel (Lehrach et al 1977), was prepared by dissolving 1g of agarose in 100 ml of 1x MOPS / EDTA buffer (10x MOPS is 0.2M 3-(N-morpholino)propanesulphonic acid, 50mM sodium acetate, 10mM EDTA pH 7). The gel was cooled to 55°C and 5.1 ml of 37% formaldehyde were added. 20 µg of total RNA in 5 µl of TE was added to 25 µl of electrophoresis sample buffer (0.75ml deionised formamide, 0.15 ml 10x MOPS, 0.24 ml of 37% formaldehyde, 0.1 ml

autoclaved water and 0.08 ml of 10% w/v bromophenol blue). The sample was heated at 65°C for 15 minutes and 1 µl of ethidium bromide (10 mg ml^{-1}) was added. Electrophoresis was carried out in 1xMOPS at 75 volts for 2 hours.

The gel was prepared for transfer to Hybond-N membranes by soaking for two 20 minute periods in 10xSSC to remove any remaining formaldehyde. The membrane was soaked in distilled water for 5 minutes and in 10xSSC for a further 5 minutes. The subsequent transfer was as for a standard Southern blot. The filter was air-dried and the RNA covalently bound by UV irradiation for one minute.

2.2.13 Labelling of DNA probes

Probes for use in hybridisation experiments were labelled with ^{32}P or digoxygenin. In both cases, the fragments were first purified by two passages through agarose gels and were recovered from the gel as described in section 2.2.10. Radioactive labelling was the more sensitive and was used in Southern genomic blots when single copy genes were sought. The digoxygenin was convenient when there was excess target DNA on the nylon filters.

2.2.13.1 Human and mouse MHC class II probes

Cloned human and mouse MHC class II probes which were used in this study are listed below.

Subregion	Probe Name	Nature of Probe	size (kbp)	Reference
Aα	p24.2	Whole gene	4.8	Davis 1984
Aβ	AβpBR	Whole gene	5.6	Malissen 1983
$Aβ_2$	Cos25.1	β2 domain	2.0	Larhammar 1985
Eα	O_2Eα	α1-CT	3.4	McNicholas 1982
Eβ	L1Eβ	β1 and β2 domains	7.2	Kronenberg 1983

Eβ2	pBR325Eβ2	β2 domain-3'UT	3.0	Denaro 1985
DPA*	pDAα13b	α1-3'UT	1.1	Trowsdale 1984
DPB	LC11	β1 and β2 domain	0.9	Trowsdale 1984
DQA	10-8	α1 and α2 domains	2.2	Trowsdale 1983
DQB*		β1 and β2 domains		Wiman 1982
DRA*	$pDRH_2$	α1 and α2 domains	1.3	Lee 1982a
DRB*	pDRβ1	whole gene	1.2	Long 1982
DZA	8bα1	α1-stop	1.7	Trowsdale 1985

* indicates a cDNA probe.

2.2.13.2 Radioactive labelling using ^{32}P

In early experiments, the DNA was labelled using the random priming method described by Feinberg and Vogelstein (1984). This procedure involved the priming of denatured DNA at many sites with random hexanucleotides and the use of the large (Klenow) fragment of DNA polymerase 1 for primer extension with the incorporation of a ^{32}P labelled dNTP. Later the Pharmacia 'Quick Prime' kit was used in which the random nucleotide mixture were 9-mers and the enzyme used was T7 polymerase. In both cases, typically, 50 ng of DNA were denatured by boiling and labelled with 50 μCi ^{32}P-dCTP (>6000 Ci/mM). The incorporation rate averaged 70% as assessed by the percentage of radioactivity which could be precipitated with 10% trichloroacetic acid. Prior to addition to the hybridisation fluid, the probe was boiled for 5 minutes and then chilled on ice.

2.2.13.3 Non-radioactive labelling using digoxygenin

DNA was labelled with digoxygenin (DIG) using a commercially available kit 'DIG DNA labelling and detection kit' (Boehringer Mannheim). In this procedure, the DNA was labelled by the random primed method as described above, but one of the dNTPs, dUTP, was labelled by attachment via a spacer arm to the steroid hapten digoxygenin.

As the Klenow catalysed reaction proceeded, one dUTP-DIG molecule was incorporated every 25-30 nucleotides.

2.2.14 Hybridisation of labelled probes to immobilised DNA or RNA

With ^{32}P labelled probes, a number of different hybridisation conditions were employed over the study period. Initially, 50% deionised formamide containing 5xSSC (20xSSC is 3m NaCl, 0.3M sodium citrate, pH 7), 5x Denhardt's solution, 0.1%SDS and 100 μg ml^{-1} denatured sheared salmon sperm DNA was used at 42^0C. For genomic Southerns, the fluid described by Church and Gilbert (1984) was used, but more recently, all hybridisations used the 'Rapid Hyb' fluid (Amersham). Hybridisations in this case were at 65^0C and could be performed in 2 hours.

Hybridisation using DIG labelled probes followed the manufacturers intructions. The hybridisation fluid was 5xSSC containing 0.1% w/v blocking reagent and 0.02% SDS. Following hybridisation, the membranes were washed to remove unbound probe. The membranes were washed at high stringency, 0.1xSSC, 0.1% SDS at 65^0C, when discrimination between MHC isotypes was required, and at low stringency, 2xSSC, 0.1% SDS, at 42^0C, when it was more important to detect sequences of lower similarity to the probe.

2.2.15 Autoradiography

Detection of ^{32}P was by autoradiography. After the final wash, the filter was blotted, wrapped in Saranwrap and autoradiographed at -70^0C using Fuji X-ray film (Genetic Research Instrumentation, Dunmow, England) for a few hours to a few days.

2.2.16 Detection of bound DIG labelled probes

DIG labelled probe which had hybridised to an immobilised target was detected immunochemically using a commercial kit (Boehringer Mannheim). Non-specific binding was blocked by incubating the membrane for 30 minutes in TN solution (100mM Tris/HCl, 150mM NaCl), containing 0.5% blocking reagent. The membranes were then transferred to a TN solution containing a 1:5000 v/v dilution of an anti-DIG alkaline phosphatase conjugate for 60 minutes. Unbound conjugate was removed by two 15 minute washes in TN buffer. The filters were then equilibrated in 100mM Tris/HCl, 150mM NaCl, 50mM $MgCl_2$, pH 9.5. The enzyme substrates were then added, 45µl NTB (75 mg ml^{-1} nitroblue tetrazolium salt in dimethylformamide) and 35µl X-phosphate (50mg ml^{-1} in 5-bromo-4-chloro-3-indolyl phosphate toludinium salt in dimethylformamide). The filters were left in the dark for a few hours to allow the purple colour to fully develope. The reaction was stopped by washing in TE.

2.2.17 Cosmid cloning of genomic DNA

Cosmids are plasmid cloning vectors which contain the bacteriophage lambda (λ) cohesive (cos) termini. The presence of the cos site enables *in vitro* packaging into phage heads (Hohn and Murray, 1978), hence hybrid cosmids can be transduced into *E. coli* with high efficiency. The λ packaging system will only package DNA molecules which are 37-52 kbp in length (Feiss et al 1977), with the optimum substrate being a linear concatamer in which the cos sites are separated by about 45 kbp of DNA. Since the cosmid vector molecules are small, 5-10 kbp, approximately 40 kbp of genomic DNA can be packaged following ligation into a cosmid vector (Collins and Hohn 1979).

There are a number of advantages to cosmid cloning. Several linked genes can be cloned on the same recombinant molecule and fewer clones need to be screened to find a gene of interest. Furthermore, cosmids replicate like plasmids. Transformed

cells grow under antibiotic selection and high yields of recombinant DNA can be obtained.

2.2.17.1 Cosmid library construction

The cosmid libraries described here were constructed using genomic DNA from three unrelated sheep. Prior to the work described here, pTL5 and pTL6 cosmid libraries were prepared and screened for ovine MHC class II genes at IAPGR by E.V.Deverson, S. Watson and by N. Huskisson. The DNA for the libraries in the cosmid vectors pTL5 and pTL6 (Lund et al 1982), came from two Suffolk rams from the flock maintained at the Institute of Animal Physiology and Genetics Research (IAPGR), Cambridge. DNA from a ram from the Finnish Landrace flock maintained at the Moredun Research Institute, Edinburgh, was used to construct the pCos8 library.

The construction of the pTL5 and pTL6 libraries is schematically described in Figure 2.1. Genomic DNA for these libraries was prepared from sperm and testis respectively using the technique described by Flavell et al (1978). The construction of the pCos8 library, which followed along similar lines to that of the earlier libraries is described in some detail here.

2.2.17.2 Preparation of genomic DNA

DNA was prepared by the method of Herrman and Frischauf (1987). 100 ml of blood from a Finnish Landrace tup was collected into 13 ml of 3.8% w/v sodium citrate. The red blood cells were lysed by addition of three volumes of lysis buffer (155mM NH_4Cl, 10mM NH_4HCO_3 and 0.1mM EDTA pH 7.4). The lysate was held on ice for 10 minutes and then centrifuged at 2000g for 10 minutes. The supernatant was removed and the white cell pellet washed with 10 ml lysis buffer, and resuspended in 4.5 ml of 25mM EDTA/75mM NaCl pH 8. 0.5 ml of 10% w/v SDS was added followed by 0.5 ml of proteinase K solution (10 mg ml^{-1} in water). The resulting

white cell lysate was incubated at 55°C overnight. The lysate was transferred to a medical flat bottle and 55 ml of phenol, freshly equilibrated in 0.1M Tris pH 8, were added. The phases were gently mixed overnight in the cold room and then separated by centrifugation at 2000g and the aqueous phase was removed to dialysis tubing

using a widebore pipette to minimise mechanical shearing of the DNA. The DNA was dialysed against TE pH 7.5 for two hours at room temperature to remove the SDS and then exhaustively at 4°C. The DNA concentration was estimated spectrophotometrically by measuring the UV absorption spectrum between 220 nm and 320 nm. Assuming a 50 μg ml^{-1} solution has an OD_{260}= 1.0, the final concentration of the DNA preparation was 70 μg ml^{-1}.

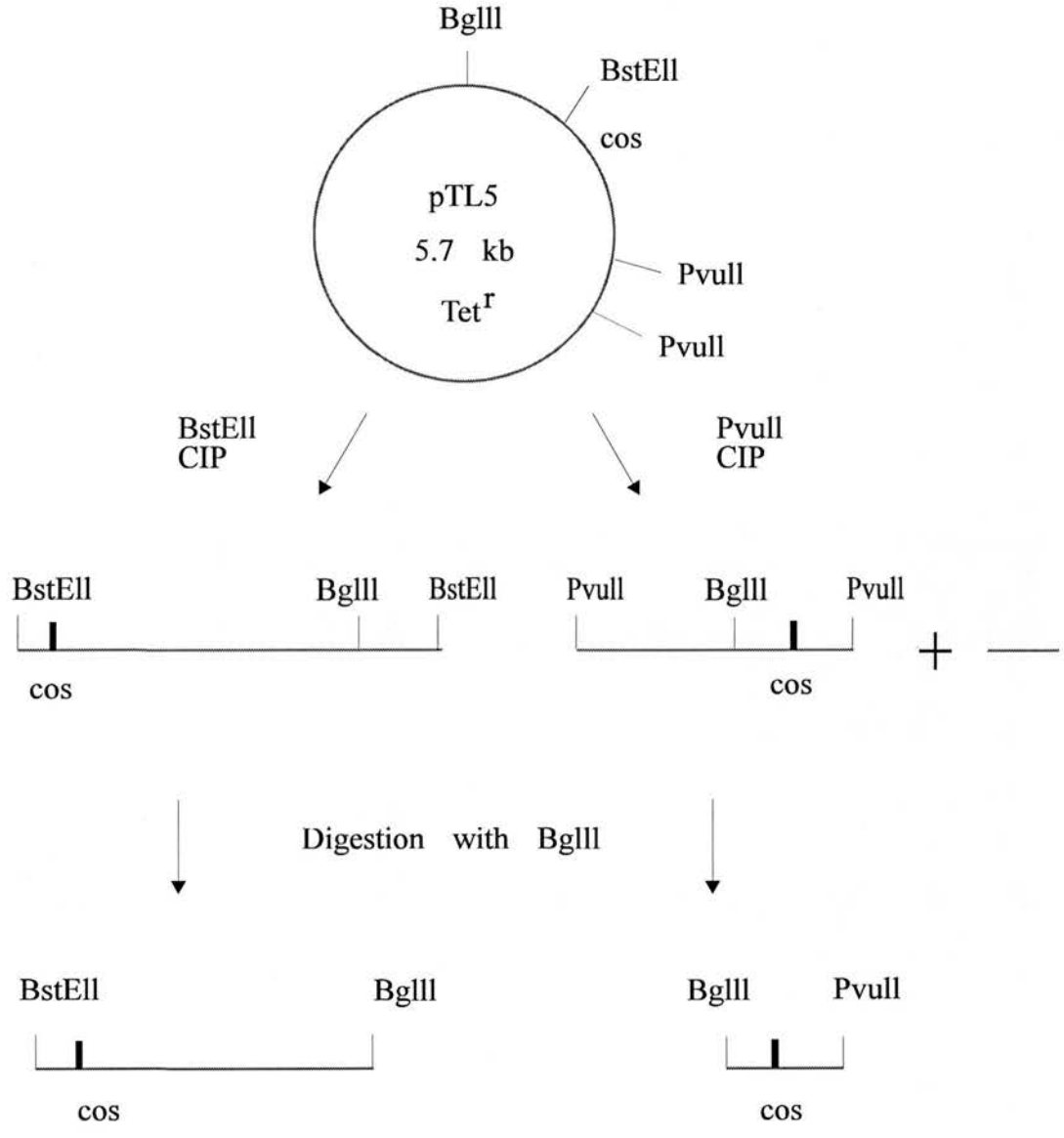
To construct a representative library, it was essential to start with DNA which was as large as possible, preferably greater than 200 kbp. Accurate sizing of DNA greater than 20 kbp is difficult using conventional agarose electrophoresis. A convenient approximation was obtained by electrophesis on a low concentration agarose gel (0.3%). The size of the DNA preparation was significantly above an uncut λ phage DNA marker (48 kbp).

2.2.17.3 Partial restriction enzyme digests of genomic DNA

The DNA was partially restricted with the enzyme Ndell which has the four base recognition sequence GATC compatible with the overhang left by the six base cutter BamHl. Small scale analytical digests were set up to titrate the enzyme. Four μg of DNA were digested with 2 μl of a 1:50 dilution of Ndell in a total reaction volume of 150 μl at 37°C. 20 μl samples were removed to tubes containing 1 μl of 0.5 M EDTA on ice after 0, 2.5, 5, 10, 20, 40 and 80 minutes. Four μl of loading buffer were added and the samples were loaded into the wells of a 20 x 20 cm 0.3% agarose gel in TAE buffer. The gel was run for 24 hours at 20 volts. The time points upto 10 minutes gave digested DNA in the size range 35-45 kb. A number of small-scale

Figure 2.1

Cosmid library construction using the vectors pTL5 and pTL6. Two aliquots of the vector were digested separately with BstEll and Pvull and de-phosphorylated (CIP= calf intestinal alkaline phosphatase). The fragments were then digested with Bglll to produce the vector arms. Tet^r= tetracycline resistance.



Gel purify vector arms and ligate to partially digested, size selected genomic DNA. Package into bacteriophage λ and transduce *E. coli* to tetracycline resistance.

stopped by the addition of EDTA. Small samples of each time point were analysed on a 0.3% agarose gel and those samples containing digested DNA in the 35-45kb range were pooled.

2.2.17.4 Size selection on NaCl gradients

Further size selection of the digested DNA was made by density gradient centrifugation. 5-25% sodium chloride gradients were prepared in two 12 ml polyallomer tubes and 200 µl samples containing 150 µg of the digested DNA were pipetted on top of each gradient. The tubes were centrifuged at 37000 rpm at 18°C for 4.5 hours in a Beckman SW40 rotor. 0.5 ml fractions were collected from the gradients. Small samples of each fraction were analysed by electrophoresis on a 0.3% agarose gel. Those fractions containing DNA in the required size range were pooled, precipitated and redissolved in 0.1xTE. The concentration of the genomic DNA was estimated by direct comparison with dilutions of a λ DNA standard. The yield was about 5 µg, ie. there was a greater than 90% loss of the starting DNA

2.2.17.5 Dephosphorylation of the genomic DNA

To prevent self-ligation of the DNA, the 5' phosphate groups were removed by digestion with calf intestinal phosphorylase (CIP) as described in section 2.2.7. The DNA was dissolved in 84 µl of 0.1xTE. A 4 µl sample was removed and kept for a pre-phosphatase test ligation. Following digestion, the DNA was extracted with phenol, precipitated and made up to 80 µl in 0.1xTE.

To assess the effectiveness of the phosphatase step, 3 µl of the pre- and post CIP DNA samples were ligated overnight at 12^0C as described above. The ligation reactions were loaded onto a 0.3% agarose gel and run overnight at 25 volts. The pre-CIP DNA self-ligated and barely left the well, whereas the post-CIP sample co-migrated with the λ DNA marker at just under 50 kbp.

2.2.17.6 Preparation of the pCos8 vector

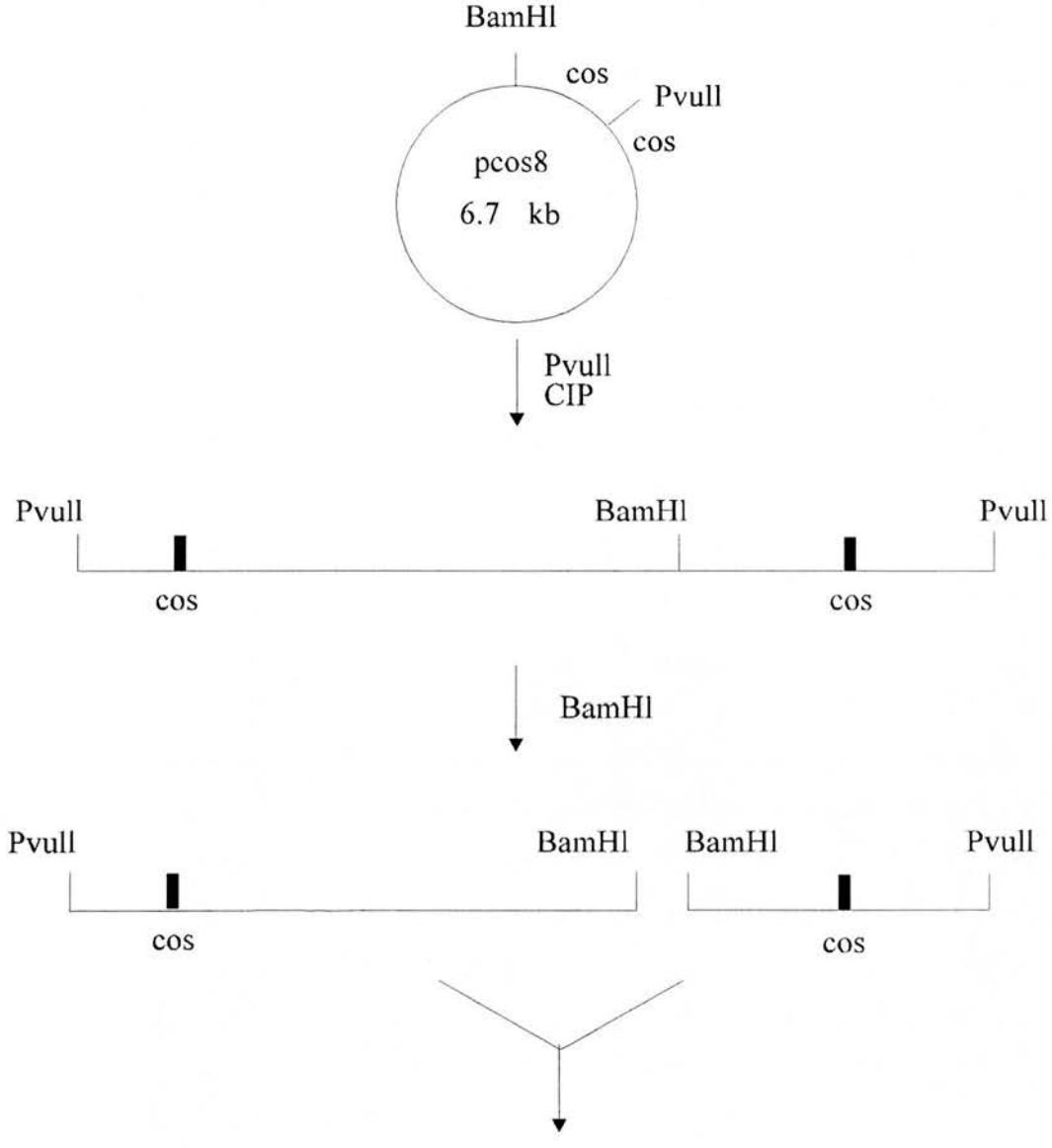
pCos8EMBL is one the family of cosmid vectors which contain two cos sites (Ehrlich et al. 1987). The cos site is derived from phage λ DNA and is formed when the 12 bp cohesive ends of the phage DNA are annealed and ligated. The steps in the production of the vector arms, each of which contains a cos site, are described below and illustrated in Figure 2.2. Purified vector DNA was produced by CsCl density gradient centrifugation as described in section 2.2.2. The vector was linearised by digestion hour at 37^0C. A further 40 units were added and the incubation left for another hour after which the DNA was extracted with phenol/chloroform and precipitated. The linear molecule was then dephosphorylated with calf intestinal phosphorylase (CIP) as described above. The DNA was again extracted with phenol, precipitated and dissolved in 20 μl of 0.1xTE.

The efficiency of both enzymic reactions was checked by test ligation of the dephosphorylated vector molecules. 1μl of the DNA was ligated overnight. JM109s were transformed with 40 ng of the test ligated DNA. Only a few colonies were obtained following plating out on L agar containing kanamycin indicating that the Pvull and CIP digests had been efficient.

The remainder of the vector DNA was digested with BamHl to produce the vector arms of about 2.1kbp and 4.4 kbp. The DNA was digested with 40 units of BamHl in a total volume of 80 μl for 1 hour at 37^0C. A further 40 units were added and the incubation left for another hour after which the DNA was extracted with phenol/chloroform, precipitated and made upto 80 μl in 0.1xTE at a final concentration of about 250 ng ml^{-1}.

Figure 2.2

Schematic diagram of the preparation of the pCos8 cosmid vector arms for subsequent ligation to genomic DNA.



Ligate to partially digested, size selected genomic DNA to form a concatamer suitable for packaging into λ phage.

2.2.17.7 Preparation of plating cells.

The *E. coli* strain 490A, was used as the host strain for the libraries. This strain carries the recA$^-$ mutation. This was confirmed prior to use by assessing the effect of UV radiation on cell growth compared with that of a wild type *E. coli* strain as described by Sambrook et al. (1989).

Plating cells were prepared as follows. 10 ml of L-broth containing 0.2% maltose and 10 mM $MgSO_4$ to induce the phage receptors, were inoculated with a single colony of 490A and shaken at 37°C during the day. 1ml was then added to 50 ml 2xTY medium containing maltose and $MgSO_4$ and the cells were grown overnight at 37°C. The flask was cooled on ice, and the cells centrifuged at 2500g for 15 minutes. The cells were resuspended in 40 ml of ice cold $MgSO_4$, re-centrifuged and resuspended in 15 ml of ice cold $MgSO_4$. Cells were prepared fresh in this way for each transduction experiment.

2.2.17.8 Test ligations and packaging of the genomic DNA into phage

Approximately 500 ng of the dephosphorylated genomic DNA were ligated into 250 ng of the BamHl / CIP'd pCos8 DNA in a reaction volume of 10 µl overnight at 16^{0}C. Packaging extracts were obtained from Amersham International, UK. A set of two packaging extract tubes were thawed on ice. 5 µl of the ligation reaction were added to the freeze-thaw extract followed by 15 µl of the sonic extract. The mixture was incubated at 22°C for two hours when 0.25 ml of SM buffer and 15µl of chloroform were added. Phage were stored at 4^{0}C.

To assess the efficiency of packaging, 5µl of the phage preparation were added to 100 µl of plating cells. The transduction was allowed to proceed for 30 minutes at 37°C. 1.5ml of 2xTY were added and the cells incubated with shaking at 37°C for 45 minutes to express the antibiotic resistance gene. The cells were transferred to an Eppendorf tube and centrifuged at 13000 rpm. The cells were resuspended in 200 µl of 2xTY and plated out on 2xTY agar plates containing kanamycin. The following

day, the number of colonies were counted. When the number of colonies per plate was in the range 200-500, the procedure was scaled up. A titred stock of phage was built up and used in a larger scale transduction to produce the library. 10 mls of plating cells were added to 500 µl of phage stock and incubated for 30 minutes at 37^0C. 150 mls of 2xTY containing kanamycin were added and the cells shaken at 37^0C for 45 minutes. The cells were centrifuged, taken up in 40 mls of 2xTY containing kanamycin and 2 mls of the suspension were spread out on each of 20, 245 x 245 mm bioassay plates (Nunc). Following incubation overnight at 37^0C, the colonies were all scraped together in a total volume of 150 ml of 2xTY, centrifuged and resuspended in 50 ml of 2xTY containing kanamycin and 15% glycerol. Aliquots of this suspension were snap frozen and stored at -70^0C.

2.2.17.9 Screening the library.

The suspension was titred by counting the number of colonies in 10 µl drops of serial log dilutions which gave a figure of $4x10^{11}$ cells ml^{-1}. The haploid sheep genome contains approximately $3x10^9$ bases. A cosmid clone contains approximately 40 kbp. Therefore one genome should be contained in $3x10^9/4x10^4 = 0.75x10^5$ clones. To be sure that most sequences were represented, three bioassay plates were each spread with $2x10^5$ cells. The cells were plated out on 20 x 20 cm Hybond-N nylon membranes. After overnight growth, the colonies were 'copied' onto a fresh sheet of Hybond-N, placed on a new 2xTY/kanamycin plate and allowed to grow at 37^0C for a few hours. The original membrane was stored at 4^0C while the copy was put through the following colony lysis procedure and screened for sheep MHC class II genes.

The membrane was placed on a filter paper soaked in 10% SDS for 3 minutes and then transferred to one soaked in 0.5M NaOH/1.5M NaCl for 5 minutes to lyse the colonies. The membrane was neutralised by being placed on a filter of 0.5M Tris/HCl pH7.5/1.5M NaCl for 5 minutes. It was then placed on a filter soaked in 2x SSC for 5 minutes before being allowed to dry in air. The dry membrane was wetted in 2xSSC

and the bacterial debris wiped off carefully with tissues. The membrane was dried and the cosmid DNA fixed to the Hybond-N by exposure to UV for 1 minute on the surface of a transilluminator.

2.2.17.10 Restriction mapping of cosmids

Detailed and accurate restriction maps are required to identify overlapping cosmid clones. Cosmids contain the cohesive-end sequences derived from bacteriophage lambda. The lambda molecule is linear, with the 12 bp cos site unannealed and unligated. In cosmids however, the cos site is ligated. Rackwitz et al (1984) mapped lambda clones by specifically labelling either the right or the left end of the clones by hybridisation to radioactively labelled oligonucleotides which were complementary to the coshesive termini. Rackwitz et al (1985) extended this type of analysis by using the enzyme lambda (λ) terminase to linearise cosmid clones. This enzyme is involved in packaging the bacteriophage DNA into the phage particles. It is specific for the cos site and provides single stranded sequences which can be hybridised to the oligonucleotide probes. Following partial digestion of the cosmid clone with selected restriction enzymes to generate all of the possible DNA fragments, those fragments which contained the cos ends were selectively labelled by hybridisation with the labelled right or left specific oligonucleotide in solution. The fragments were then resolved on agarose gels and, following autoradiography, the sizes of the labelled fragments were estimated. The fargment size was equal to distance in kbp of the various retriction sites from the ends of the clone and thus a map could be constructed.

2.2.17.11 Linearisation of cosmid DNA using λ terminase

λ terminase was obtained from Amersham, cat. no. RPN 1720). Digests were set up as follows

cosmid DNA	2 μg	DTT 50mM	2 μl
buffer A	2 μl	λ terminase	2 μl

buffer B	2 μl	water	to 20 μl
ATP 100mM	1 μl		

Buffer A :150mM Tris/HCl, 22.5mM $MgCl_2$, 7.5mM EDTA, pH8.

Buffer B : 6mM Tris/HCl, 18mM $MgCl_2$, 30mM spermidine and 60mM putrescine pH7.4.

The reaction mixture was incubated for 30 minutes at 20^0C and then heated to 65^0C for 3 minutes.

2.2.17.12 Generation of partial restriction digests of the linearised cosmids.

Partial digests were contrived by digesting the DNA in two dilutions of enzyme, 1:20 and 1:50, and by taking samples of each digest at 5, 10 and 15 minutes. The reactions were stopped by the addition of 1 μl of 0.5M EDTA. Four μg of DNA were digested in a total volume of 20 μl with 1 μl of the enzyme dilutions.

2.2.17.13 5' end-labelling of oligonucleotide probes

Each of the two oligos ON-L and ON-R were diluted to a concentration of 1 pM/μl and their 5'-OH ends were labelled with ^{32}P using the enzyme polynucleotide kinase. The ^{32}P was derived from [γ-^{32}P]ATP. The reaction conditions were as follows :

ON-L or ON-R	4.5 μl
[γ-^{32}P]ATP 5000Ci/mM	3 μl (30 μCi)
10x kinase buffer*	1.5 μl
Polynucleotide kinase	1 μl (3 units)
water	5 μl

* 10x kinase buffer : 1M Tris/HCl, 100mM $MgCl_2$, 70mM DTT, pH8.0

The reaction was allowed to proceed at 37^0C for one hour before the enzyme was killed by heating to 65^0C for 3 minutes. For use in hybridisation experiments, 2 μl of

each probe were diluted in a mixture of 125 µl TE8, 75 µl gel loading buffer and 50 µl 1M NaCl. The remaining probe was stored for upto two weeks at -20^{0}C.

2.2.17.14 Hybridisation of labelled probes to the partially digested cosmids.

The partial digest of the cosmid was divided between two tubes and 5 µl of labelled ON-L added to one and 5 µl of ON-R to the other. The tubes were placed at 70^{0}C for 3 minutes to melt the cos ends and then at 42^{0}C for 30 minutes to allow the oligos to anneal. The samples were loaded into the wells of a 20x20 cm agarose gel and the labelled fragments were separated by electrophoresis overnight at 25 volts in Loening E buffer (36mM Tris, 1mM EDTA, 34mM NaH_2PO_4 pH 7.6). Molecular weight markers were prepared by digesting λ DNA with a number of different enzymes. The markers were then treated in the same way as the cosmids. At the end of the run, the gel was dried under vacuum onto a sheet of Whatman DE81 paper which binds unincorporated ^{32}P-ATP, and autoradiographed overnight.

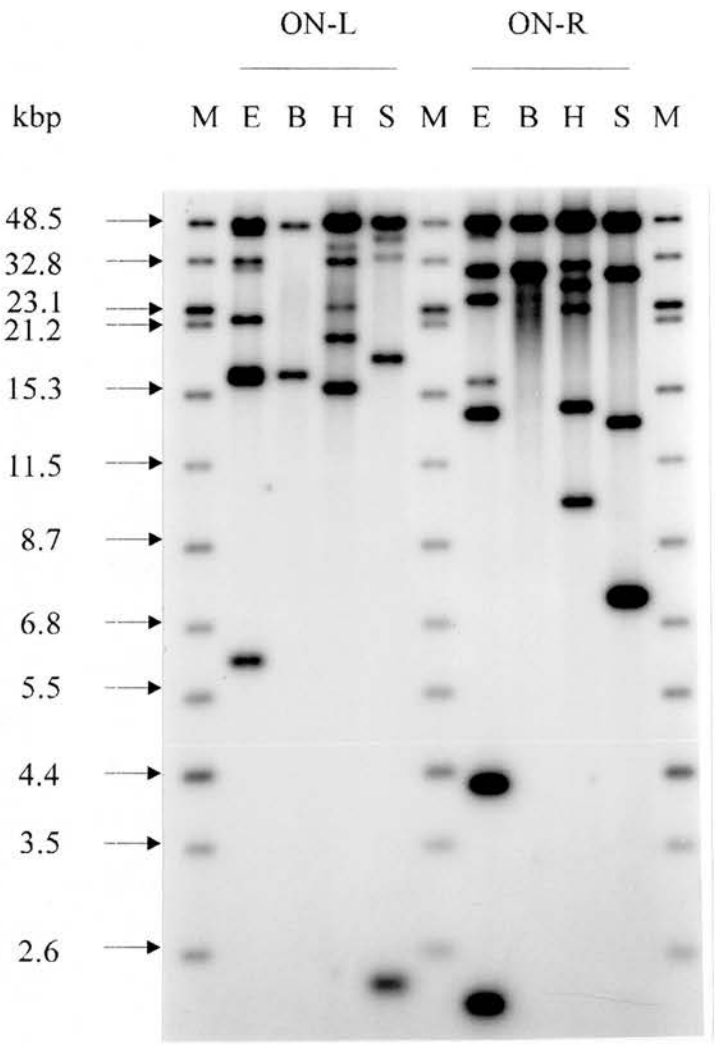
2.2.17.15 Restriction maps

Approximately forty clones were mapped in this way using the five enzymes EcoR1, BamH1, Hindlll, Sac1 and Sma1 and a typical result is shown in Figure 2.3. Data from the autoradiograph was read into a computer program (Mapgel) using a sonic digitiser and the fragment sizes and hence the map distances were calculated from the standards using a cubic spline algorithm written by J Coadwell of IAPGR, Babraham.

The sizes of all of the fragments resulting from complete digests of the cosmids with the same enzymes were also calculated and used in combination with the mapping data to construct the final maps. Southern blots of the complete digests of the cosmids were hybridised to various class ll probes and hence those fragments which contained the various sheep class ll genes could be identified on the maps.

Figure 2.3

Autoradiograph of a cosmid mapping gel produced as described in sections 2.3.17.10 to 2.3.17.15. ON-L and ON-R refer to the oligonucleotides complementary to the left and right coshesive ends of λ phage. The tracks are labelled M, molecular weight markers (kbp); E, EcoRl; B, BamHl; H, Hindlll and S, Sacl.



2.2.18 DNA sequencing

All DNA sequencing described here used the chain-termination method first described by Sanger et al 1977. This method involves the synthesis of a DNA strand by a DNA polymerase. Synthesis is initiated at the point where an oligonucleotide, usually a 17-mer, has been annealed to a single stranded DNA template. The reaction is terminated by the incorporation into the growing DNA strand of the 2',3'-dideoxynucleoside 5'-triphosphates (ddNTPs) which lack a 3'-OH group necessary for further polymerisation. For each DNA template to be sequenced, four enzyme catalysed reactions are set up. Each reaction tube contains all of the normal nucleotides dATP, dTTP, dCTP and dGTP plus a radioactively labelled nucleotide, together with one of each of the four ddNTPs. As the DNA chains are synthesised they become radioactively labelled, and some are terminated prematurely by the incorporation of the ddNTP analogue. If the relative concentrations of the normal and the dideoxy nucleotides are correct, then each possible chain length of DNA will be produced. The DNA chains are then separated on large, denaturing polyacrylamide gels and the bands corresponding to chains which terminate at specific bases are visualised by autoradiography.

2.2.18.1 Preparation of templates for sequencing.

In the initial stages of this work all of the subcloning of the cosmid DNA was done in the vector pBS+. Double stranded DNA was prepared as described in section 2.2.2 and denatured and sequenced as decribed below. An advantage of pBS+ is that it contains sites for the universal sequencing primer and the M13 reverse primer flanking the multiple cloning site allowing the insert to be sequenced in both directions. Furthermore, larger inserts can be sequenced following the production of a set of nested deletions as described in section 2.2.18.4.

However, a disadvantage which was encountered, was the occurrence of compressions due to secondary structures in the DNA which made the reading of the

sequence difficult. These difficulties were ameliorated by the use of the deaza-dGTP analogue which acts by reducing the strength of the intramolecular hydrogen bonding.

By using specific oligonucleotides (17-mers) as sequencing primers, sequence data could be obtained directly from the 40 kbp cosmid following alkaline denaturation. Without doubt however, the best templates were single stranded DNA prepared by subcloning into the filamentous bacteriophage M13mp18/19 or by rescue of single stranded DNA, as M13K07 phage, following subcloning into the phagemid Bluescript SK (+/-) (Stratagene).

2.2.18.2 Preparation of single stranded template from M13mp18/19 clones.

M13 is a filamentous phage specific to *E. coli* strains which express the sex pili encoded by an F factor. When the bacterial cell is infected by single-stranded viral DNA, the cellular enzymes convert the single-stranded form to the double stranded (RF) form. Initially, replication of the viral genome converts the newly synthesised single stranded DNA into the RF form. However, as certain viral gene products build up in the cell, the production of the RF form is switched off, ssDNA begins to accumulate and phage particles are secreted by the cell. The host cell is not lysed by the phage, but does grow rather more slowly and hence is recognised as 'plaques' on a lawn of normally growing cells (Messing 1983).

To produce the M13mp18/19 series of cloning vectors, a short segment of the *E. coli* genome was introduced into the region between the viral II and IV genes (Yannich-Perron et al 1985). This segment of *E. coli* DNA codes for the first 146 amino acids of β-galactosidase (lacZ). The F' plasmid carried by appropriate host cells carries a defective lacZ gene which codes for an inactive enzyme which lacks amino acids 11-41. By α-complementation, the NH_2-terminal of the protein produced in cells containing an M13 vector associates with the defective cellular protein to produce an active enzyme. Messing and colleagues built into the *E. coli* segment a multiple cloning site containing 13 restriction sites which can be used to insert foreign DNA.

This had no effect on α-complementation and meant that recombinant molecules could be distinguished from non-recombinant on the basis that those cells which contain a non-recombinant can enzymically turn the chromogenic substrate X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) blue following enzyme induction by IPTG (isopropylthio-β-D-galactoside). Hence recombinant plaques are white while non-recombinant plaques are blue.

The multiple cloning site of the RF form of M13mp18/19 was cut with an appropriate restriction enzyme(s) and ligated to a fragment of sheep DNA. An aliquot of the ligation mixture was used transform competent JM109 cells. 200 μl of an overnight culture of JM109, grown from a single colony picked from a minimum agar plate, were added to the transformed bacterial cells together with 40 μl of X-gal (20 mg ml^{-1} in dimethylformamide) and 40 μl of IPTG (2% in water). 3 mls of top agar at 48^0C were added, mixed and quickly poured onto an L-agar plate to be incubated overnight at 37^0C.

Next day a single white plaque was picked into 2 mls of L-broth and shaken at 37^0C for 5 hours. The culture was transferred to an Eppendorf tube and the cells centrifuged at 13000 rpm for 10 minutes. The supernatant was transferred to a fresh tube and respun. 1.2 mls of the supernatant was transferred to a fresh tube and 300 μl of 20% PEG 6000, 2.5M NaCl solution were added to precipitate the phage particles. After 15 minutes at room temperature the tube was centrifuged at 13000 rpm for 10 minutes. The supernatant was carefully removed and the phage particles were resuspended in 100 μl of TE. 50 μl of phenol were added and the tube vortexed for 30 seconds and then left at room temperature for 15 minutes before being re-vortexed. The phases were separated by centrifugation for 5 minutes at 13000 rpm. 10 μl of 4M LiCl and 250 μl of ethanol were added to the supernatant to precipitate the phage DNA. The DNA was pelleted at 13000 rpm, washed with 70% ethanol and dissolved in TE.

2.2.18.3 Rescue of single stranded DNA from the phagemid Bluescript using the filamentous helper phage M13K07.

Phagemids are vectors which combine the advantages of both plasmids and filamentous bacteriophages. The Bluescript series of phagemids (Short et al 1988) carry an origin of replication, an ampicillin resistance gene and the region between genes II and IV of the phage f1 which regulates gene expression and viral DNA synthesis. The helper phage M13K07 carries a mutation in this intergenic region, an origin of replication and the kanamycin resistance gene. Viral gene II product interacts with this intergenic region to initiate rolling cycle replication of single stranded viral DNA. However, when cells carrying the phagemid are infected with M13K07, the product of gene II of the incoming virus interacts preferentially with the wild type intergenic region carried by the phagemid, resulting in the accumulation of phagemid ssDNA. This DNA is then packaged into phage particles in normal way.

Bluescript subclones in JM109, were streaked out on L-agar plates containing ampicillin and incubated overnight at 37°C. Single colonies were picked into 2 mls of 2xTY broth and 2 x 10^7 pfu of the M13K07 phage added. The culture was shaken at 37°C for 1.5 hours. Kanamycin was added to a final concentration of 70 $\mu g\ ml^{-1}$ and the incubation at 37°C continued overnight. The preparation of ssDNA from the phage was completed as in section 2.2.18.2.

2.2.18.4 Generation of a set of nested deletions for DNA sequencing.

DNA fragments up to 500bp in length are conveniently sequenced by cloning into plasmids such as Bluescribe or into M13 vectors. Fragments larger than 500 bp can be sequenced by the generation of subclones of the original clone which form a set of nested deletions in which the primer binding site in the vector is steadily brought closer to the sequence of interest.

The Erase-a-Base system (Promega Corp. Madison, Wisconsin, USA) is based on the method of Henikoff (1987). Exonuclease lll is used to digest the insert DNA in a vector such as Bluescribe, from a 5' protruding or blunt ended restriction site produced by digestion with enzymes such as BamHl. The primer binding site is protected from Exo lll digestion by a four base 3' overhanging restriction site produced by digestion with enzymes such as Sphl.

Exo lll digestion proceeds at approximately 450 bp per minute at 37^0C, and so deletions of a given length can be contrived by removing timed aliquots from the reaction. The single-stranded tails are removed by digestion with S1 nuclease. The low pH and zinc ion concentration of the S1 buffer inhibit further Exo lll digestion. Following heat inactivation of the S1 nuclease, the ends of the DNA strands are repaired by Klenow DNA polymerase. Subsequent ligation and transformation of *E. coli* yields a library of subclones in which the insert has been deleted to a greater or lesser extent.

The required buffers and reagents are as follows

10x Exolll buffer	:	660 mM Tris-HCl, pH 8.0 6.6 mM $MgCl_2$
7.4x S1 buffer	:	0.3 M Potassium acetate pH 4.6, 2.5 M NaCl, 10 mM $ZnSO_4$, 50% glycerol.
S1 stop buffer	:	0.3 M Tris base, 0.05 M EDTA
Klenow buffer	:	20 mM Tris-HCl pH 8.0, 100 mM $MgCl_2$
dNTPs	:	0.125 mM each of dATP, dCTP, dGTP and dTTP
10x ligase buffer	:	500mM Tris-HCl, pH 7.6, 100 mM $MgCl_2$, 10 mM ATP
Ligase mix	:	790µl deionised water
		100µl 10x ligase buffer
		100µl 50% polyethylene glycol
		10µl 100mM DTT
		5 units T4 DNA ligase

10μg of plasmid DNA were digested with 20 units of Sph1 to produce the 3' overhang, and with 20 units of BamH1 to produce the 5' overhang, in a 50μl reaction volume at 37°C for 2 hours. The DNA was extracted with phenol/chloroform and precipitated by the addition of 0.5 volume of 7.5M ammonium acetate and four volumes of ethanol and then dissolved in 60μl of 1x Exo lll buffer. 500 units of Exo lll endonuclease were added and the mixture incubated at 37°C. 2.5 μl samples were removed at 15 second intervals and added to 7.5μl of S1 nuclease mix on ice. This mixture was then incubated at room temperature for 30 minutes. 1μl of S1 stop buffer was added and the S1 nuclease was inactivated by incubation at 70°C for 10 minutes. 2 μl samples were taken and run on 0.8% agarose gel to check the extent of the deletions.

1μl of Klenow mix and 1μl of the dNTP mix were added to each time point and the tubes incubated at 37°C for 5 minutes. 40 μl of ligase mix were then added and the tubes incubated at room temperature for 30 minutes. 10μl samples from each time point were used to transform JM109s to ampicillin resistance in the usual way. Colonies were picked and DNA prepared as usual.

2.2.18.5 Annealing of primer to template.

The sequencing primers used were the M13 reverse primer, the universal primer or specific 17-mers synthesised by Oswel DNA, Edinburgh. 1-2 μg of single stranded DNA in a volume of 10 μl were placed in a tube, to which was added, 2μl of annealing buffer (200mM Tris HCl, pH7.5, 100mM $MgCl_2$, 250mM NaCl), and 0.5pM of primer in 2μl water. This mixture was heated to 65°C for 2 minutes and then allowed to cool to approximately 30°C over a period of 30 minutes.

Double stranded plasmid DNA was denatured in alkaline solution prior to the annealing reaction. Approximately 4 μg of plasmid DNA in a volume of 8 μl of water was denatured by the addition of 2 μl of 2M NaOH. After 10 minutes, 3μl of 3M

sodium acetate pH 4.8, 7µl of water and 60µl of ethanol were added and the DNA precipitated at -70°C. After centrifugation, the pellet was washed in 70% ethanol and dissolved in 10µl of water ready for annealing.

2.2.18.6 Sequencing Reactions

The sequencing reactions were performed using two commercial kits, Pharmacia ^{T7}Sequencing kit, cat. no. 27-1682-01 and the Sequenase version 2.0 kit with 7-deaza dGTP (USB cat. no. 70750).

To the primed template the following reagents were added and in the order stated. 3µl of labelling mix (1.375 µM each of dCTP, dGTP and dTTP in 333mM NaCl), 1µl of ^{35}S-dATP (10µCi) and 2µl of the DNA polymerase diluted 1:5 in enzyme dilution buffer which contained glycerol, bovine serum albumin and dithiothreitol in Tris/HCl pH 7.5. Following incubation at room temperature for 5 minutes, 4.5 µl aliquots were added to each of four tubes which contained 2.5 µl of one of the four termination mixes (containing one of the ddNTPs). After 5 minutes incubation at 37°C, the reactions were stopped by the addition of 5 µl of stop solution (deionised formamide containing 10mM EDTA, xylene cyanol and bromophenol blue dyes).

2.2.18.7 Sequencing gel electrophoresis

The apparatus used was the IBI model STS 45 sequencing tank. A 6% polyacrylamide gel was prepared by dissolving 42 g of ultrapure urea in 20 ml of 30% w/v acrylamide / 0.8% w/v bis-acrylamide stock solution, 10ml 10xTBE buffer and water to 100 ml. The polymerisation reaction was initiated by adding 0.8 ml of 10% w/v ammonium persulphate and 40µl of N,N,N,N-tetramethylethelenediamine (TEMED). The gel was formed between two 43 cm x 38 cm glass plates separated by a 0.4 mm spacer. A sharkstooth comb was used to form the sample loading wells. The running buffer was 1xTBE and the gel was pre-run for 1 hour at 80W to heat to 50°C. The samples were denatured at 80°C for 2 minutes immediately prior to loading. 2.5

µl aliquots of each sample were pipetted into neighbouring wells in the order G, A, T and C. Electrophoresis proceeded until the bromophenol blue dye reached the bottom of the gel. Further samples were loaded as necessary. At the end of the run, the plates were separated and the gel fixed in a mixture of 10% v/v acetic acid and 10% v/v methanol for 20 minutes. The gel was removed from the fix and transferred to a supporting sheet of Whatman 3MM filter paper, covered with Saranwrap and dried under vacuum at 80^0C. The dry gel was exposed to X-ray film overnight.

The nucleotide sequence was read from the pattern of bands on the autoradiograph. Computer analysis of the sequences made use of the Genetics Computer Group package (Devereux et al 1984) on the SEQNET facility of the Science and Engineering Research Council, Daresbury Laboratory, Warrington, England.

2.3 Analysis of sheep MHC gene expression

The expression of sheep MHC class II genes was assessed at two levels. Firstly by analysing specific mRNA production using reverse transcription-polymerse chain reaction (RT-PCR) and Northern blotting (section 2.2.12). Secondly, by DNA-mediated gene transfer of cosmid DNA into the mouse L-cell to assess cell surface expression of a class II glycoprotein.

2.3.1 Preparation of total cellular RNA

The tissue was snap frozen in liquid nitrogen and a sample ground up to a fine powder in a mortar and pestle in liquid nitrogen. The tissues were homogenised in 5ml of RNA extraction buffer (GIT, 23.6g of guanidine isothiocyanate, 5ml of 250mM sodium citrate pH 7, 2.5ml of 10% w/v N-lauroylsarcosine, 0.36 µl β-mercaptoethanol, 47ml of water) using the technique described by Chirgwin et al (1979). 0.5ml of 2M sodium acetate were added followed by 5ml of water-saturated phenol and 1ml of chloroform:isoamyl alcohol (24:1) and the mixture shaken vigorously for 5 minutes. After 10 minutes on ice, the solution was centrifuged at

12000 rpm for 30 minutes at 4°C. The aqueous phase which contained the RNA, was removed and the RNA precipitated by the addition of 5ml of isopropanol. The precipitate was pelleted and dissolved in 0.5ml GIT extraction buffer. 0.5ml of isopropanol was added. The precipitate was pelleted, washed in 70% ethanol and resuspended in 100 µl of autoclaved water.

2.3.2 Analysis of gene expression by reverse transcription polymerase chain reaction (RT-PCR)

First strand cDNA was prepared from total cellular RNA using AMV reverse transcriptase. Specific sequences within the cDNA were then amplified using the polymerase chain reaction.

2.3.2.1 cDNA synthesis

A commercial kit was used for cDNA synthesis (cDNA Cycle Kit, Invitrogen). 5 µg of total cellular RNA were placed in a tube in a total volume of 11.5 µl of sterile distilled water. 1 µl of primer was added (oligo dT or random hexamers or a specific internal primer, a 17-mer) to a final concentration of 1 µM. The sample was heated to 65°C for 10 minutes to remove the secondary structure and then placed at room temperature for 2 minutes. The following were then added in the order indicated, 1 µl of RNase inhibitor, 4µl of 5x RT buffer (250mM TRIS-HCl, 40mM $MgCl_2$, 150mM KCl, 5mM DTT, pH 8.5), 1µl 100mM dNTPs, 1µl 80mM sodium pyrophosphate, 0.5µl (5 units) of AMV reverse transcriptase. The mixture was carefully mixed, the tube centrifuged and placed at 42°C for 1 hour. The tube was incubated at 95°C for 2 minutes to denature the RNA-cDNA hybrids and then placed on ice. The sample was then ready for PCR amplification.

2.3.2.2 Polymerase chain reaction (PCR)

The reaction was set up in a total volume of 20 µl as follows, 2 µl, approximately 100ng cDNA, 2 µl 10xPCR buffer (100mM Tris/HCl, 15mM $MgCl_2$, 500mM KCl, pH 8.3), 2 µl of each primer at 10µM, 2 µl dNTPs each at 2 mM, 9 µl water and 1 unit Taq polymerase (Boehringer). 50 µl of mineral oil was added and the tubes placed in a Hybaid Omnigene thermal cycler.

PCR reactions involved 25 cycles of denaturation at 95^0C for 2 minutes, followed by 2 minutes at an annealing temperature calculated to be 5^0C below the melting temperature (T_m) of the two oligonucleotide primers, followed by 3 minutes extension at 72^0C.

Samples of the amplified products were run out on agarose gels and their identity proved by Southern blotting and hybridisation to specific probes.

2.3.3 Expression of sheep MHC class II genes following DNA-mediated gene transfer

The mouse fibroblast cell line Ltk^- (Spandidos and Wilkie, 1983), does not express endogenous mouse MHC class II genes. It is therefore a suitable cell line in which to study the expression of foreign MHC class II genes. Cloned genes can be expressed in these cells under the influence of their own promotors following DNA-mediated gene transfer, using the calcium phosphate technique described by Graham and van der Eb (1973) as modified by Wigler (1979). The method depends upon the endocytosis of a microprecipitate formed between the genes to be expressed and calcium phosphate. In a small percentage of cells, endocytosed DNA is incorporated into the mouse genome and expressed.

The L-cell is deficient in endogenous thymidine kinase (tk^-) and will not survive in the presence of the nucleotide metabolic inhibitor aminopterin. This provides a means of

cell selection. The herpes simplex virus tk gene has been cloned (Lang et al 1985, pTK1), and is included in the calcium phosphate / DNA co-precipitate. Cells expressing the pTK1 gene and other co-transfected genes can be selected in medium containing aminopterin, providing that thymidine and hypoxanthine are also present (HAT selection, Szybalska and Szybalski 1962).

Those cells that survived HAT selection, were monitored for surface expression of sheep class II expression in an indirect fluorescence assay (IFA). The first stage antibody was a mouse anti-sheep class II mab, and the second, a fluorescein isothiocyanate (FITC) conjugated rabbit anti-mouse IgG. The cells were analysed by FACScan and positive cells selected on a FACS IV cell sorter. FACS-selected cells were expanded in culture and re-sorted to produce a stable mouse L-cell line expressing a sheep MHC class II isotype.

2.3.3.1 Cells and media.

The L-cells were cultured in 25 cm^3 tissue culture flasks (Nunc, Gibco-BRL, Paisley) containing 8 ml Dulbecco's modified Eagles medium (DMEM, Gibco-BRL, cat. No. 04101965) supplemented with 10% v/v heat-inactivated fetal bovine serum (FBS, Northumberland Biologicals, Cramlington), 2mM glutamine, 100 IU ml^{-1} penicillin, and 100 μg ml^{-1} streptomycin. Cells were maintained in this medium and split 1:20 twice weekly. Selection medium was as above, but also contained 1.36g l^{-1} hypoxanthine, 0.176g l^{-1} aminopterin, and 0.388g l^{-1} thymidine.

2.3.3.2 Monoclonal antibodies.

Monoclonal antibodies (mabs) that recognise monomorphic determinants of sheep MHC class II glycoproteins have been described previously. The various mabs, together with their proposed specificities are listed in table 2.1 which is taken from Ballingall et al. (1995)

Table 2.1 Monoclonal antibodies specific for sheep MHC class ll molecules.

Monoclonal antibody	Ig isotype	Alpha / Beta chain specificity	Previously proposed subtype specificity
SBUII 28.1, 37.68	IgG1	Conformational	DR
SBUII 38.27	IgG1	Conformational	DQ
SBUII 42.20	IgG1	Alpha	DR
SBUII 49.1	IgG2a	Beta	Pan-beta
SW73.2 (Rat)	IgG2b	Beta	Pan-beta
VPM 1	IgM	Conformational	ND
VPM 4, 46	IgG2a	Beta	ND
VPM 45	IgG2a	Beta	DQ
VPM 36	IgG1	Alpha	DQ
VPM 38, 47, 54	IgG1	Alpha	DR
VPM 58, 59	IgG1	Conformational	DR
VPM 37, 57, 43	IgG1	Beta	DR
VPM 16, 40, 41, 44	IgG1	Beta	DQ

The SBUII mabs have been described previously (Puri et al.. 1985, 1987, Puri and Brandon 1987). Their proposed specificity was based on N-terminal sequence analysis. The VPM mAbs have also been described previously (Hopkins et al 1986, Dutia et al. 1990, 1993, 1994). The sub-type specificity of VPM 36, 37, 38 and 54 was determined by N-terminal sequence analysis (Dutia et al. 1993, 1994). The specificity of the remainder of the VPM mAbs was determined by ELISA of mAb supernatants against affinity purified ovine *DQ* and *DR* antigens (Dutia et al. 1993, 1994 and unpublished observations). SW73.2 reacts with both *DQ* and *DR* antigen (Hopkins et al. 1986, Dutia et al. 1993). ND specificity has not been determined.

2.3.3.3 DNA-mediated gene-transfer and selection of transfected cells

6 μg of cosmid DNA were added to 100 ng of pTK1 plasmid and 10μg of high molecular weight mouse embryo carrier DNA, in a total volume of 500 μl. 60 μl of 2M $CaCl_2$ were added. The $CaCl_2$/DNA solution was added dropwise with vigorous agitation to 0.5ml of 2xHBS (made up by mixing 5ml of 1M HEPES, 0.75ml Na_2HPO_4 and 28.8 ml of 1M NaCl made up to 100ml with distilled water and the pH adjusted to 7.0 with HCl). The tubes were left for 30 minutes for the calcium phosphate microprecipitate to form. The microprecipitate was then mixed with $5x10^5$ exponentially growing L-cells. After an incubation period of approximately 15 hours at 37°C the cells were washed in warm phosphate-buffered saline. They were then left to recover in fresh medium at 37°C for 24 hours before HAT selection. Selection medium was changed every 3 days and surviving colonies allowed to develop for 2-3 weeks. Cells transfected with pTK1 alone were used as negative controls for antibody assays.

2.3.3.4 Assay for MHC class II expression.

$5x10^5$ L-cells were incubated for 40 minutes at 4°C in a cocktail of mouse anti-sheep MHC class II monoclonal antibodies (SBU 28.1, 37.68, 38.27, 42.20 and 49.1) using saturating amounts (50 μl tissue culture supernatant). After three washes in cold Earle's balanced salt solution* containing 3% foetal bovine serum (FBS) and 0.1% (w/v) sodium azide, the cells were incubated in 50μl of a 1:50 dilution of fluorescein isothiocyanate (FITC) conjugated rabbit anti-mouse IgG (Dakopatts). After a further three washes, 100 μl of a 200 μg ml^{-1} solution of propidium iodide were added to distinguish between the highly fluorescent dead cells and the cells expressing sheep MHC class II. The cells were then analysed by FACScan (Becton Dickinson) and positive class II expressing cells selected on a FACS IV cell sorter. FACS-selected cells were expanded in culture and sorted a further three times until more than 98% of the cells expressed MHC class II glycoproteins.

*Earle's balanced salt solution was made as follows.

Solution A: 2.6g of $CaCl_2$ in 200ml water.

Solution B: 2g of $MgSO_4$ in 200ml water.

Solution C: 68g of NaCl, 4g of KCl, 1.4g of Na_2HPO_4 in 600ml of water. The solutions were mixed together and filter sterilised.

Working EBS: 1xEBS supplemented with 5ml 1M HEPES buffer pH 7.2-7.4, 0.1% w/v sodium azide, 3% serum to 500ml with sterile distilled water.

Phosphate buffered saline: (PBS) NaCl 60g l^{-1}, KCl 4g l^{-1}, Na_2HPO_4 1.48g l^{-1}, KH_2PO_4 4g l^{-1}.

Chapter 3

The sheep MHC class II *DQ* sub-region

3.1 Introduction

The MHC is a highly polymorphic multigene family found in all vertebrates studied thus far (Klein, 1986). Much of the polymorphism in the class ll genes of the HLA, is located in the *DQ* sub-region, at both the *DQA1* and *DQB1* loci (Marsh and Bodmer 1993). The proteins encoded by these different allelic forms act as restriction elements in the presentation of processed antigen to the cells of the immune system (Schwartz 1985). In humans, a duplication of the *DQ* genes occurred approximately 20 million years ago (Trowsdale et al. 1985), resulting in the formation of the *DQA2* and *DQB2* genes (formerly called *DX*α and *DX*β*).* Unlike the *DQ1* genes, the *DQ2* genes are relatively non-polymorphic (Jonsson et al. 1987, Berdoz et al. 1989). Furthermore, although the *DQ2* genes show no characteristics of pseudogenes (Auffray et al. 1987), they are not expressed in any of the different tissues tested (Collins et al. 1984, Auffray et al. 1987) and no *DQ2* product has been described.

In the mouse, the genes orthologous to *HLA-DQ1*, are designated *I-A*α and *I-A*β (Steinmetz et al. 1982a). They also are highly polymorphic (Klein 1986) and act as restriction elements (Benacerraf et al. 1978, Allen et al. 1987). However, in the mouse there has been no gene duplication (Steinmetz et al. 1982a,b, 1986).

Early studies of class ll genes of the sheep MHC used HLA gene probes to analyse sheep genomic DNA (Millot et al. 1984, Chardon et al. 1985, Scott et al. 1987). Genomic bacteriophage λ clones were isolated which, although it was not formally demonstrated, were considered to contain sheep orthologues of the *HLA-DQA* and *DQB* genes (Scott et al. 1987).

In cattle, RFLP studies had shown the cattle *DQA* and *DQB* genes to be highly polymorphic (Andersson et al. 1986a,b), and that the number of *DQ* genes varies between MHC haplotypes (Andersson and Rask 1988). In humans, the number of *DQ* genes is constant regardless of MHC haplotype (Bohme et al. 1985)

At the start of this study therefore, little was known of the detailed structure of the *DQ* sub-region of the sheep MHC, no nucleotide sequences of the genes were available and it was not known which genes were expressed. In this chapter, the *DQ* region of the sheep is described in terms of a restriction map of the region, constructed from overlapping cosmid clones. Nucleotide sequences of some alleles at the different loci are presented. Expression of sheep *DQ* class ll molecules at the surface of the mouse L cell is also described.

3.2 Cosmid clones containing *DQ*-like genes

A preliminary analysis of all of the cosmid clones obtained from the pTL5 and pTL6 libraries, ie. from sheep 1 and sheep 2, was performed by probing dot-blots of the cloned DNA with individual HLA class ll probes (Deverson et al. 1991). The differential hybridisation of the clones to the various probes showed that cosmids 4, 22.1, 39.1, 2 and 9.2 contained *A* genes which were *DQ*-like in character. Furthermore, cosmids 4, 22.1 and 39.1 also hybridised to HLA class ll *B* gene probes.

Restriction maps of these clones for the enzymes EcoRl, BamHl, Hindlll, Sacl and Smal were produced and compared (Figure 3.1). It was clear that cosmids 4, 39.1 and 22.1 overlapped, with only one site which was unique to cosmid 4 indicated by a closed box, and four which were unique to cosmid 39.1 indicated by open circles. Southern blots of restriction digests of the cloned DNA hybridised to human *A* and *B* genes gave the positions of the genes and showed that this cluster contained one *B* gene which was flanked by two *A* genes. Cosmids 9.2 and 2 came from sheep 1 and 2 respectively but had identical restriction maps (Figure 3.1). They contained a single *A* gene apparently unrelated to the others. A further clone, cosmid 62, contained a single *B* gene. The map of cosmid 62 (Figure 3.1), showed that, in common with cosmids 22.1 and 39.1, it contained only two Smal sites. The fragments bounded by these sites were similar in that they hybridised only to exon 2-specific *B* gene probes.

Figure 3.1

Restriction map of the sheep MHC class ll DQ sub-region.

See text for details. E=EcoRl, B=BamHl, H=Hindlll, Sa=Sacl, Sm=Smal.

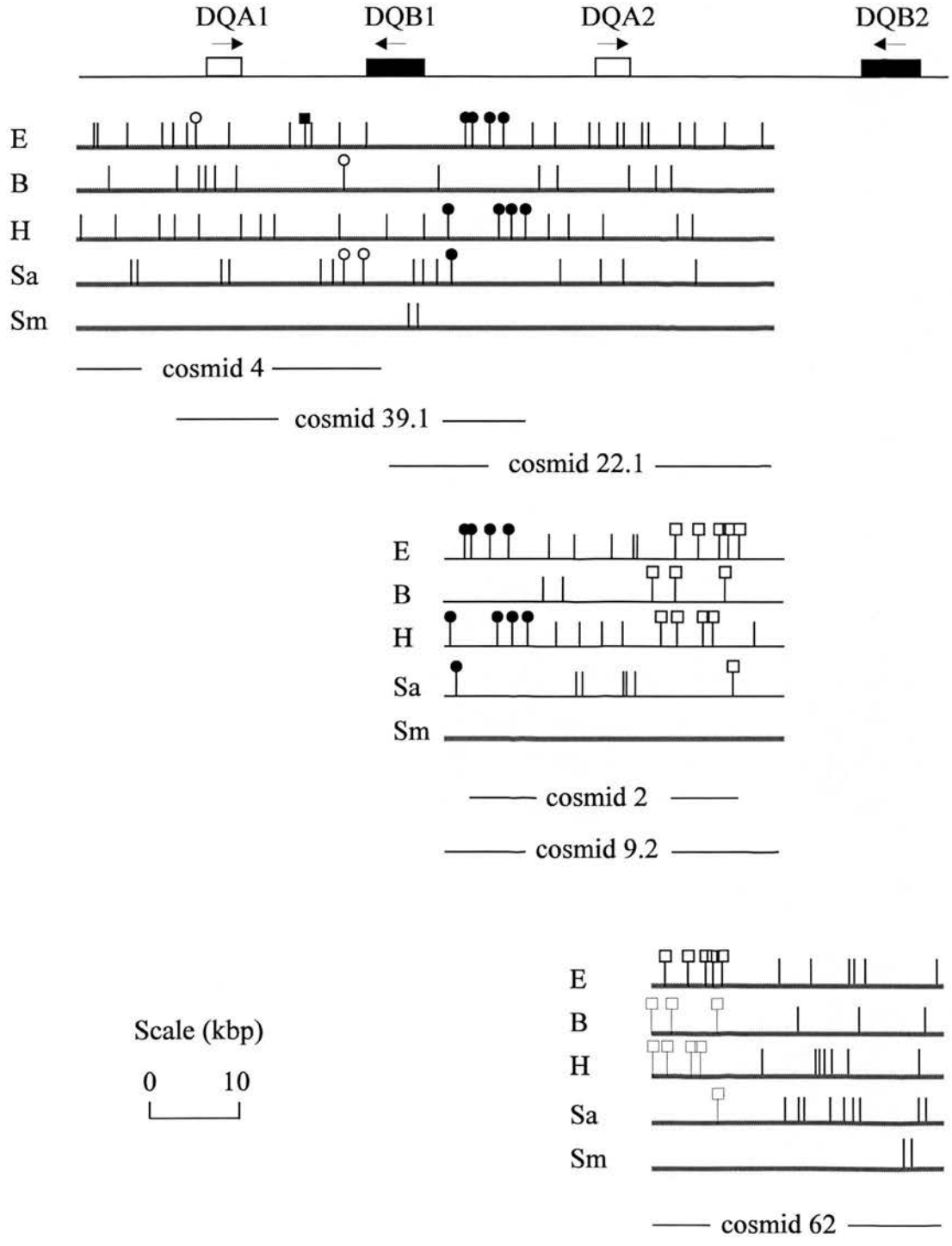


Figure 3.2

Hindlll restriction digests of the cosmids from sheep 3 which contain a *DQA/DQB* gene pair. The positions of the λ Hindlll molecular weight markers are marked with arrows.

Track No.	Cosmid	Track No	Cosmid
M	λ Hindlll	4	307
1	313	5	381
2	314	6	343
3	341	M	λ Hindlll

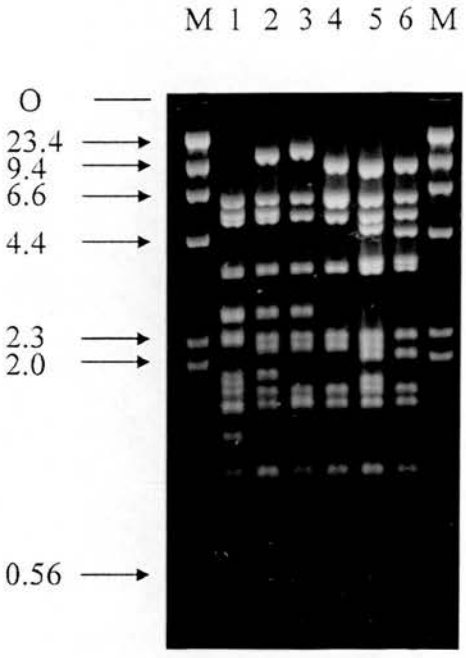


Figure 3.3

Restriction maps of overlapping cosmids 341, 313, 314 (cluster 1), and 381, 307 and 343 (cluster 2), for the enzymes EcoRl, BamHl, Hindlll, Sacl and Smal. Within a cluster, sites were a perfect match. Between clusters, those sites marked with an asterisk matched.

Transcriptional direction is indicated by arrows. The position of the *A* genes is indicated by an open box while that of the *B* genes is indicated by a filled box.

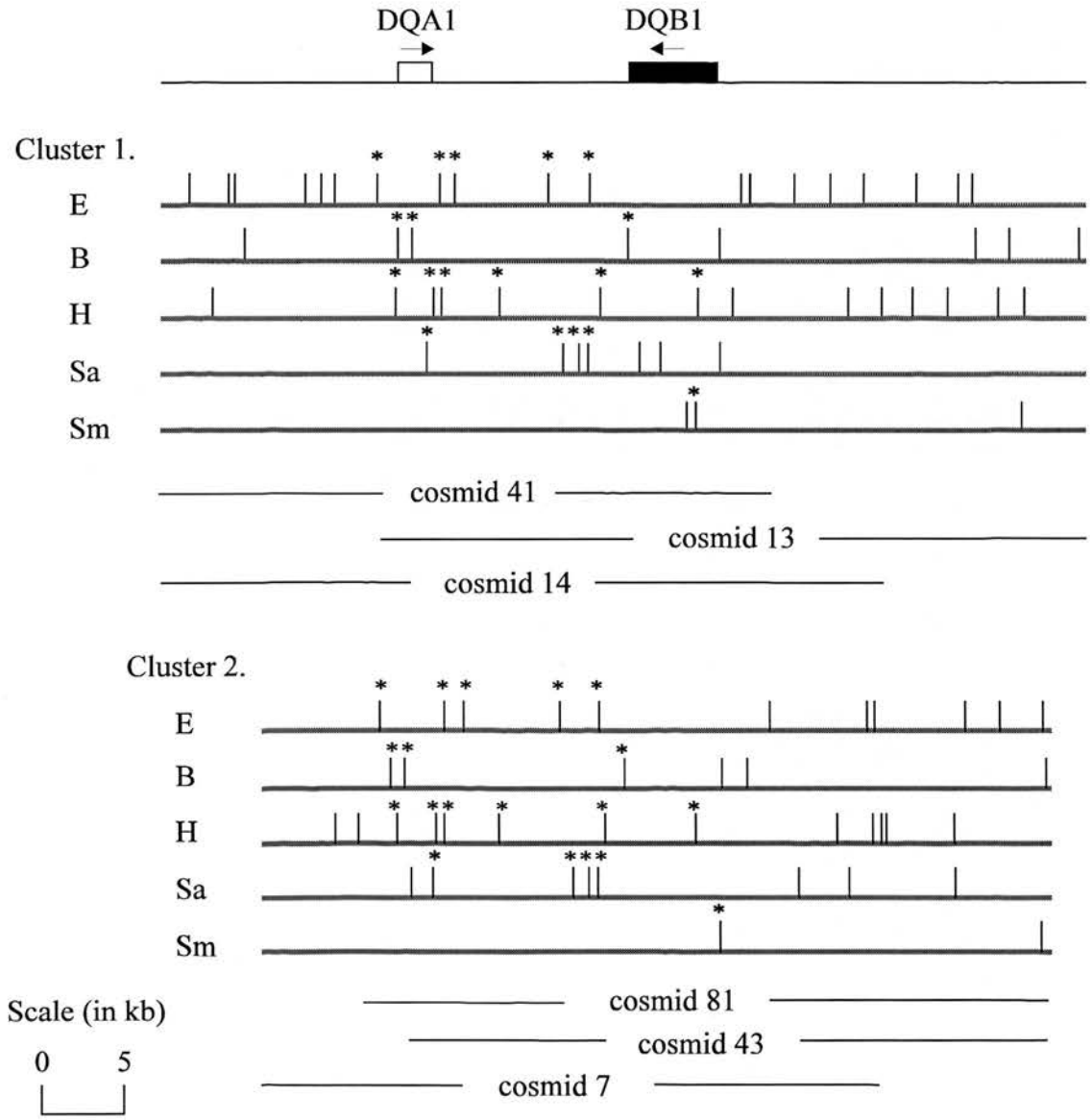
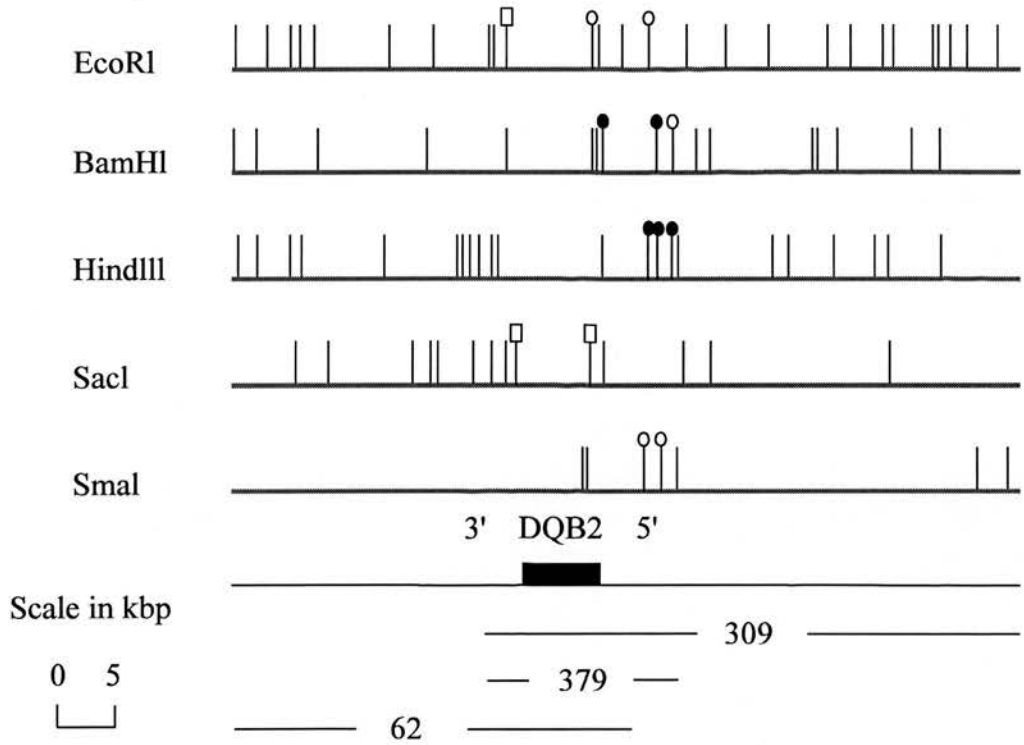


Figure 3.4

Restriction maps for cosmids 309 and 379 from sheep 3 and cosmid 62 from sheep 2. Open circles represent sites unique to cosmid 379, closed circles represent sites unique to cosmid 309, and squares represent sites unique to cosmid 62.



Parallel studies on the expression of sheep class II genes following DNA-mediated transfection of cosmid DNA from clones obtained from the pTL5 and pTL6 libraries into mouse L cells, succeeded in demonstrating expression of the *DR* isotype, but failed to show expression of any of the *DQ*-like genes (Ballingall, 1991, Ballingall et al. 1992). It was decided therefore to construct a new cosmid library to obtain more *DQ* clones. The construction of this library was described in section 2.2.17. Screening of the library yielded a number of class II positive clones. HindIII restriction digests of six cosmids, numbers 307, 313, 314, 341, 343 and 381, demonstrated the close similarity between these clones (Figure 3.2). The cosmids hybridised to both an HLA class II *A* and a *B* gene probe. The six cosmids could be arranged into two clusters of three overlapping clones based on their restriction maps (Figure 3.3). Within a cluster, restriction sites for all five enzymes matched perfectly. Between clusters, the sites within the areas containing the two genes matched well, as indicated by the asterisks in Figure 3.3. Two further clones from the pCos8 library, cosmids 309 and 379, contained single *B* genes as shown in Figure 3.4. Cosmid 379 was unusual in having a shorter than usual insert of approximately 20 kbp (compared to 40-45 kbp), despite of the physical and biological selection applied during the library construction. Cosmids 379 and 309 shared a short area of overlap with cosmid 62.

It was considered that the only way to define the affinities of all of these genes fully was by nucleotide sequence analysis.

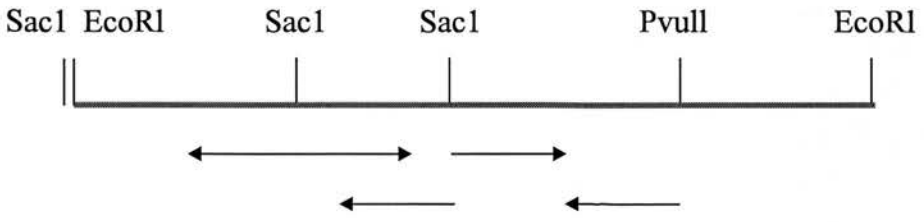
3.3 The nucleotide and amino acid sequences of the cosmid *A* genes.

Hybridisation of Southern blots of cosmid 2/9.2 digested with EcoRI, BamHI, HindIII, SacI and SmaI, with an *HLA-DQA* probe indicated that the smallest unique hybridising fragment was a 2.1 kb EcoRI fragment. This fragment was sub-cloned into the vector pBS+. A restriction map of the sub-clone was produced (Figure 3.5). The two small SacI fragments were cloned (one making use of the SacI site in pBS+), as was the 600 bp SacI/PvuII fragment. These subclones were sequenced. Database searches using the FASTA sequence comparison program, revealed that this was

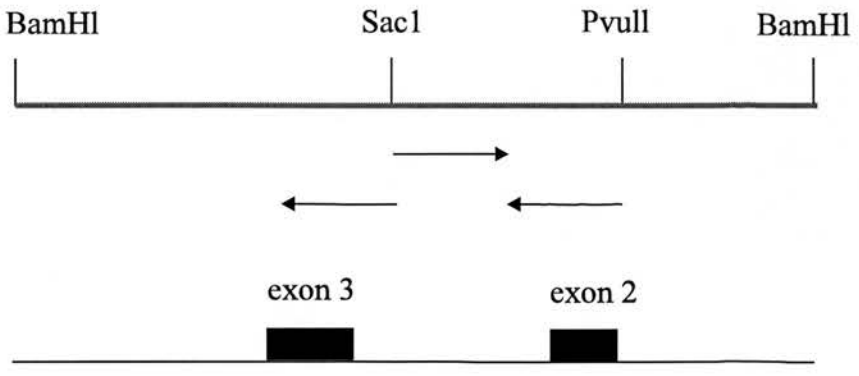
Figure 3.5

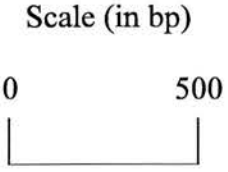
Restriction maps and sequencing strategies for the BS+ subclones of the DQA genes in cosmids 2 and 4. Note that the leftmost SacI site in subclone 2 is in the vector.

A. 2.1 kb EcoRI subclone from cosmid 2.



B. 2.1kb BamHI subclone from cosmid 4





indeed a *DQA*-like gene. The Sac1/Pvull fragment contained the second exon of the gene. A similar analysis of cosmid 4, identified a 2.1 kbp BamHl fragment which hybridised to *DQA* probes. The map of this fragment is also shown in Figure 3.5. This clone also contained a 600 bp Sacl/Pvull fragment which when sequenced was shown to contain the second exon of a *DQA* gene.

At this point the assumption was made that many, if not all, sheep *DQA* genes contain a Pvull site at the beginning of exon 2, and have a Sacl site positioned 600 bp downstream. Therefore, to circumvent unnecessary sub-cloning and to obtain the fragment which contained the more interesting and polymorphic second exon, the remaining cosmids 9.2, 39.1, 22.1, 381 and 314 were doubly digested with Sacl and Pvull and following electrophoresis the band at 600 bp was excised from an agarose gel. The fragment was cloned into M13mp18 and 19 and sequenced.

The nucleotide sequences of the second exons of all of the *A* genes are given in Figure 3.6, together with those of a genomic clone and two cDNA clones taken from the literature. The amino acid translations of these same sequences are shown in Figure 3.7. As is described in section 3.8 below, the only *DQ* genes which were expressed following transfection into mouse L-cells were those in clusters 1 and 2 (Figure 3.3). Therefore, in Figures 3.6 and 3.7, the other sequences are compared to the expressed *DQA* gene in cosmid 314. EMBL Acc. No. m33305 is a sheep *DQA* gene obtained on a bacteriophage clone by Scott et al. 1991a, while m93430, m93431 and m93433 are sequences obtained from sheep cDNA clones (Fabb et al. 1993). The *DQA* gene in cluster 1 (pCos8 cosmid 314) was identical to the other sheep *DQA* gene described by Scott et al. (1991a) acc. no. m33304. The sequences of m33305 and m93433 are identical and so only one is included.

The sequences clearly fall into two distinct groups based on their relative homologies (Figure 3.6 and Table 3.1). The first group contained the two expressed *DQ* genes which we have therefore designated *DQ1* by analogy with the *HLA-DQ1* genes which represent the expressed *DQ* locus in the human MHC. There were only 4 nucleotide

Legend to Figure 3.6

The nucleotide sequences of the second exons of *Ovar-DQA1* and *DQA2* genes from the cosmids and from published sequences. The sequences are identified as indicated in the table below. The splice junctions are underlined. Dashes indicate identity.

1. cosmid 314	Expressed	DQA1	6. cosmid 2/9.2		DQA2
2. cosmid 381	Expressed	DQA1	7. Scott et al. m33305		DQA2
3. cosmid 39.1		DQA1	8. cDNA Fabb et al. m93431		DQA2
4. cosmid 4		DQA1	9. cosmid 22.1		DQA2
5. cDNA Fabb et al. m93430		DQA1			

Figure 3.6

```
1  CAGCTGACCACATTGGCACCTATGGCGTAAACATCTACCAAACATATGGTCCCTCTGGCT 60
2  ------------------T-------------G--------------------------
3  ---------------C-G--------A----TG-------CT------------------
4  ---------------C-G--------A----TG-------CT------------------
5  --------------------------------TG--------------------------
6  -----------T-----T--------A--C-TG--------T-TC--------------C
7  -----------T-----T--------AC-G-G---------T-TC--------------C
8     --------G------T--------C-G--T--------T-TC--------------C
9  -----------G---------------C-G-AT--------T-TC-----------A--G

1  ACTATACCCATGAATTTGATGGAGATGAAGAGTTCTACGTGGACCTGGAAAAGAGGGAGA 120
2  ------------------------------------------------------------
3  ------------------------------------------------------------
4  ------------------------------------------------------------
5  ---T-------------------------C-----------------------C------
6  -G--C--------------------C--GCT---T--T----------GG----A-----
7  -G--C-----G--------------C--GCT---T--T----------GG----A-----
8  -G--C-----C-----------G--C--GCT---T--T----------GG----A-----
9  -G--C-----G--------C-A---C--GCT-C-T--T----------G----AA----

1  CTGTCTGGCGTCTGCCTGAGTTTAGTAAATTTACAAGTTTTGACCCTCAGGGTGCACTGA 180
2  ------------------------------------------------------------
3  ---------A--------T-------C------G--------------------------
4  ------------------T-------C------G---G-----T---------------
5  -----------------AT-------------GG-GA-------------TT--------
6  -A--------G------AT------CC-G---G--G----------A---C-----G---
7  --------A-G------AT------CC-G---G--G-------T--A-------------
8  ----------G------AT----G--G----C--------------G--A----------
9  ----------G------AT----G-CC-G---G--G-----C--AT---A-T--------

1  GAAACATAGCTACGGTGAAACATAATTTGGAGATCTTGATTCAAAGGTCCAACTCTACTG 240
2  ---------------C------------------------------T-------------
3  --------------------------------G--------------------------
4  -------C---GT--G------GTC-----------------------------------
5  --------CG----C----G------------C--A-A----------------------
6  TTC-AT-------ATC---G--C--C-----TTA-A---C-A---AC-------T---CC
7  -TG-A--------A-CA-----A--C-----T---C---C-A--C-C-------T---CC
8  -TG-A--------A-C------C--C-----T---A-----A--C-T-------T---CC
9  -T-----------A-A------C--C-----TG--A---C-A--T---A-----T---CC

1  CTGCTACCAACAGTATGTGTTCAC
2  ---------------C--------
3  ---------------A--------
4  ---------------A------G-
5  ------------
6  A---C-T----G--------C---
7  ------T---TG---A----C---
8  ---T--T---TG
9  CA-T--T---TG---A----C---
```

differences between the sequences of the *DQA1* genes in clusters 1 and 2. The nucleotide sequences within the group designated *DQA1* were 88-98% similar to one another and they presumably represent alleles at this locus. The nucleotide sequences of the *A* genes within the second group, *DQA2*, were 81-90% similar to one another, but only 73-81% similar to those in *DQA1*.

Table 3.1 also shows the identity at the amino acid level, within and between the two groups of *DQA* genes. The 4 nucleotide differences between the two expressed alleles translated to 3 amino acid changes. Within the *DQA1* group, the percentage amino acid identity varied from 77% to 96%. Corresponding Figures for the *DQA2* genes were 74% to 82%. Between groups the sequences shared 52-70% identity.

39 nucleotide positions out of the 243 in the second exon of the *DQA1* sequences reported here are variable. When this was translated into amino acids, 22 of the 81(27%) in the α1 domain of the protein chain were polymorphic. As has been shown in other species, this polymorphism was grouped into three hypervariable regions corresponding to the amino acid residues 9-18, 47-56 and 67-82. This distribution is of course of functional importance, as these positions correspond to regions of the class ll molecule which contribute to the structure of the antigen binding groove. The *DQA2* group of sequences were more variable than the *DQA1* group, with 51 nucleotide changes translating to 34 (42%) amino acid changes. Some of the non-polymorphic residues were specific to either *DQA1* or *DQA2*. This was true at positions 19, 42, 74, 78, 82 and 86. In the case of amino acid position 19, tyrosine was specific not only to the sheep *DQA1* but also occurred in the *HLA-DQA1**0101 allele, while the sheep *DQA2* group and the *HLA-DQA2* allele (Jonsson et al 1987) were both represented by a histidine residue. At six polymorphic sites, the two groups contained amino acids which were specific to that group, namely positions 8, 14, 24, 34, 64, 69.

Figure 3.7

Multiple alignment of the amino acid translations of the second exons of the *DQA* genes whose nucleotide sequences are given in figure 3.6. The first five sequences have been designated *DQA1* and the second four *DQA2*. M93430 is a sheep *DQA1* cDNA and m93431 is a sheep *DQA2* cDNA (Fabb et al. 1993). M33305 is from a genomic clone (Scott et al. 1991). The human *DQA1*0101* allele is shown for comparison as is the human *DQB2* gene (Jonsson et al. 1987). Dashes indicate identity.

```
              10        20        30        40        50        60        70        80
cos314      DHIGIYGVNIYQTYGPSGYYTHEFDGDEEFYVDLEKRETVWRLPEFSKFTSFDPQGALRNIATVKHNLEILIQRSNSTAATN
cos381      ---------V-------------------------------------------------A---------M--------
cos39.1     ---AA--I-V-HS-----------------------------H--V--Q-R-------------------V------------
cos4        ---AA--I-V-HS-------------------------------V--Q-RR-----------VG-QS---------------
m93430      ----T----V---------F--------Q---------------M----GD----F-----R-L-----LI-----------

cos2/9.2    --F-S--IHV--SH----Q---------L-----G-K-------M--Q-AG----R-VIQL--S----DYMTKH--F-H-I-
m33305      --FGS--TE---SH----Q--Q------L-----G-K-------M--Q-AG-------SE---A-Q--D--TK---F-P-I-
m93431      --V----ADF--SH----Q---------L--V--G-K-------M-GE----------SE---A----D-M-K---F-PVI-
cos22.1     --V-T--AEF--SH---SE--Q---E--LL------K-------M-GQ-AG-HI-V--S----E----DVMTKWY-F-PVI-

HLADQ*0101  --VASC---L--F-----Q----------------RK--A--W------GG---------M-VA----N-M-K-Y-------
HLADQA2     --VAS----F--SH----Q----------------TK----Q--M----I-----S----M-VG--T--FMMRQ--------
```

Table 3.1. The percent identity between the nucleotide and amino acid sequences of the second exons from the *DQA* genes in the various cosmids and published ovine *DQA1* (Fabb et al. 1993) and *DQA2* genes (Scott et al 1991a, Fabb et al. 1993). Note that the sequences of the *DQA1* gene of Scott et al. (1991a) was identical to that of cosmid 314, and the *DQA2* of Scott et al. (1991a) and Fabb et al. (1993) were identical.

	DQA1 (Fabb)	314	381	4	39.1	DQA2 (Scott)	2/9.2	DQA2 (Fabb)	22.1
					amino acids				
DQA1 (Fabb et al.)		85	85	77	80	63	63	63	57
314	92		96	83	86	67	63	70	57
381	91	98		83	85	66	65	70	57
4	93	92	91		92	67	68	62	52
39.1	88	95	93	95		65	65	64	57
DQA2 (Scott et al.)	76	78	78	77	78		82	81	81
2/9.2	77	75	75	73	77	90		75	81
DQA2 (Fabb et al.)	75	81	81	76	77	90	85		74
22.1	75	76	75	75	77	87	88	88	
					nucleotides				

3.4 The nucleotide and amino acid sequences of the *B* genes.

As indicated above, the two SmaI sites in cosmids 39.1, 22.1, 62, 314, 379 and 309 bounded the second exons of each of these genes. The SmaI fragments were sub-cloned and sequenced. The nucleotide sequences of the second exons from all of these genes are shown in Figure 3.8 together with that of another sheep *DQB* gene (Scott et al. 1991b). Although their restriction sites did not match exactly (Figure 3.4), the nucleotide sequences of the second exons from cosmids 379 and 309 were identical. In contrast to the comparatively large differences between the nucleotide sequences of the two groups of *DQA* genes, all of the *B* genes were between 90% and 94% similar to one another and it was therefore difficult to decide whether or not they were alleles at one locus or whether they should be assigned to different loci. The *B* genes adjacent to the *DQA1* and *DQB2* genes shown in Figures 3.1 and 3.3 have been designated *DQB1* and *DQB2* on the basis of this proximity. The 21 nucleotide differences between the expressed *DQB1* genes in clusters 1 and 2 resulted in 14 amino acid changes with only three silent substitutions. The polymorphic nucleotides grouped together into five hypervariable regions as defined for human and mouse (Lundberg and McDevitt 1992). Of the 270 nucleotide sites, 64 were polymorphic (24%).

The amino acid translations of the second exons are given in Figure 3.9 together with those of two other sheep (Scott et al. 1991b, Fabb et al. 1993), and two cattle *DQB* genes, *Bota-DQB1*2.0* (Xu et al. 1991) and *Bota-DQB2*3* (Sigurdardottir et al. 1992b). In both Figures 3.8 and 3.9.the sequences are compared with that of one of the expressed sheep *DQB1* genes. The sequences contain the conserved cysteine residues involved in the intramolecular disulphide bridge formation in the membrane-distal domain of the protein, and a putative glycosylation site at position 19 (Figure 3.9). The overall identity between the sheep genes listed varied from 75-84%. 36 (41%) of the amino acid positions out of 88 were polymorphic. There are clearly blocks of highly conserved amino acid motifs with intervening hypervariable regions.

Figure 3.8

Multiple alignment of the nucleotide sequences of the *Ovar-DQB* genes in cosmids 314, 381, 39.1 (*DQB1*), and cosmids 62 and 309/379 (*DQB2*). Gaps have been inserted to optimise the alignment. Dashes indicate identity. The splice sites are underlined. Exon sequences are in capital letters.

1. Cosmid 314
2. Cosmid 381
3. Cosmid 39.1
4. Scott et al. 1991b DQB
5. Cosmid 62
6. Cosmid 379/309

```
1 cccgggttctcagcgggaggcgcagggcggggctggagcggaacaggggctgagcg      60
2 ---------g------------------cc----------g------c-------acgg
3 ---------g------------------c--------------------------gcgg
4 --t------g--     -----------c--------a-c----------c-----gcgg
5 ------g--g------------------c--------------------------gcgg
6 ---------t------------------c---------------------c-----gcgg


1                          ccccgctggcg gcggccggcacgtccccatctgg 120
2          gtttcaggtttaggga----------- ----t------------------
3       cgggttgccggtttcggga----------- -----------c-----------
4       cgggttgccggtttcggga ----------c-----------c-----c----
5 cggtttcgggtttcgggtttcggga----------- -----------c-----c-----
6        cggttgccggtttcggga----------- -----------c-----c-----


1 ccgcgcggccccgcgtggggctgtggggctgagcctgaccgagcggctgtctccccgcag 180
2 ---A-------------------------------------------------------
3 -------------------------------------------------------------
4 ---A--- -G--C------------T---------------------------------
5 ---A---C--------------------C-------------------------------
6 ---A------------------C-------------------------------------

            HVR1
1 AGGATTTCGTGTTTCTGTTTATGGGCCAGTGTTACTTCACCAACGGGACGGAGCGGGTGC 240
2 -----------CAC-A------A-------------------------------------
3 -----------G-C-A-----A-----T-------------------------------
4 ------------AC-A------A-----C-------------------------------
5 -----------CAC-A-----A-----G-----G---------------------------
6 -----------CAC-A-----A-T---T-------------------------------

    HVR2
1 GGCTCGTGACCAGATACATCTACAACCAGGAAGAGCACCTGCGCTTCGACAGCGACTGGG 300
2 ------------------------------G---TT-G--------------------
3 --AGT----A----------------------G---T--G--------------------
4 ---AT-------------------------G---TT-G--------------------
5 ---AT---------------------G---G---T--GGCG-----------------
6 --TA----------------------G---G---T--GC--------------------
```

```
                          HVR3
1 GCGAGTACCGGGCGGTGACGCCGCCGGGGCAGCGGCAAGCCGAGTACTTCAACAGCCAGA 360
2 A------------------C----T-----G------C---------------------
3 A------------------C----T-----G--C-G-C----------GG----------
4 A----C-------------C----T-----G--C-G-C----------GG----------
5 A----------------------T-----G----AGC----------GG---------G
6 A---------C-G-------------------------------------GG----------

  HVR4
1 AGGACATCCTGGAGCGGACCGGGGCCGAGGCGGACACGGTGTGCAGACACAACTACCAGG 420
2 ---------------A---GCA-----C--T---------------A------------
3 ---------A-------GTGC-------------------------------------A
4 --------A----------GC---------T----------------------------
5 -----T---------A---GC---------T-----G--------------------G--
6 -----T-------------GC---------T----------------A---------G--

  HVR5
1 TGGAAGCCGCCTTCACCTGGCAGCGGCGAGgtgagtgccggccgccctccgcgggcccgc 480
2 ------------------------------------------------------------
3 AC-GCT-AT-ACAT-------------------------------------------------
4 ----------------------------------------    ---------G----
5 --T-T---C-----------------------------------------ccc-gg
6 --T-T---C-----------------------------------------ccc-gg


1 cctccacccgccaggactccgcaccgaagggactgagtcctccggggcggggtccccaga 540
2 -------------------------------
3 -- tac----------------g-a-g----g----------------------------
4 ------------------------------ -------- --------tc-- --------


1 cccacggatgggacagagagccgctgagggacaggggccgagggcacaacgtgagggtg 600
2         ----------------------------------------g--------c-
3 ------------------------------------a-------ag--g----------
4 -- -----------------g----------------c--ag------g--------c-


1 ggggtgggcatcgagtttggcaacccggg                                629
2 ---    -------a-------g---t--
3 ------------------
4 ---    -------a------- ---t--
```

Figure 3.9

Multiple alignment of the amino acid translations of the second exons from the *DQB* genes in the cosmid clones 314, 381, 39.1, 62 and 309 together with translations of two sheep and two cattle *DQB* genes taken from the database.

Dashes indicate identity. x indicates residues within the antigen binding site. *Spaces* indicate deletions. * indicate sequence not available.

β1 domain

```
    10        20        30        40        50        60        70        80        90
    xx  x              x x      xx                  x   x     x  xx  x           x
1 DFVFLFMGQCYFTNGTERVRLVTRYIYNQEEHLRFDSDWGEYRAVTPPGQRQAEYFNSQKDILERTGAEADTVCRHNYQVEAAFTWQRR
2 ---HQ-I-----------------------FV------D-------L-R-H------------Q-Q-AV-----N------------
3 ---VQ-K-L-----------S-N--------YV------D-------L-RPD---W---------VR-----------N-LITSL---
4 ---YQ-I-H-----------H----------FV------D-H-----L-RPD---W------M---R--V------------------
5 ---HQ-K-R-----------H-------R--YA------D-------L-R-S---W---E-F--Q-R--V-R------R-Y-P------
6 ---HQ-KCL-----------Y-------R--YA------D---G-----------W-----F----R--V-----N--R-Y-P------
7 ---HQ-K-L--------------------R--YA------------- L-R-----W------T--VR--V---------A-LITSL---
8 ---YQ-K-L-----------Y----------YV-----V--------L-RPP--HW-----F--Q-R--V-S-------L-LITSL---
9 ***-Q---------------Y----------FA------D-------L-RPD--LW----------R--V-R---N----D-P-*****
```

1 cosmid 314
2 cosmid 381
3 cosmid 39.1
4 *Ovar-DQB1* (Scott et al. 1991)
5 cosmid 62
6 cosmid 309
7 *Ovar-DQB1* mRNA (Fabb et al. 1993)
8 *Bota-DQB1*2.0* (Xu et al.1994)
9 *Bota-DQB2*3* (Xu etal 1994)

Three amino acid residues were specific for the *DQB2* genes in cosmids 62 and 309, namely the R, Y, and P residues at positions 84, 86 and 88. Three of the sites within the antigen binding site of the protein molecule were monomorphic in sheep and in the cattle sequences shown viz. positions 11, 30 and 78. In some cases the sheep genes were more similar to those of the cattle genes than they were to other sheep genes. For example, the *DQA1* gene in cosmid 39.1 was more similar to that of the *Bota-DQB1*2.0* than it was to the gene in cosmid 314.

3.5 Phylogenetic analysis of the sheep DQA and DQB genes

A phylogenetic analysis of the amino acid translations of the second exons of various class ll *A* and *B* genes was peformed as described in Appendix 1. As shown in Appendix 1, Figure A1.1, this analysis placed all of the *DQA1* genes together with *HLA-DQA1*0101* and *HLA-DQA2* on one statistically significant branch. The sheep *DQA2* genes grouped together on a different branch.

As shown in Figure A1.2, the branch containing the two sheep genes designated *DQB2,* was not significant being supported by only 83% of bootstrap trials. The *DQA1* genes in cosmids 314 and 381, and the sequence of Scott et al. (1991b) were more similar to the bovine sequence *BotaDQB2*3* than to the sheep sequence in cosmid 39.1.

3.6 A restriction map of the *DQ* sub-region

The high degree of sequence homology between the *A* genes in cosmids 4, 39.1, 314 and 381 on the one hand, and between the *A* genes in cosmids 2/9.2 and 22.1 on the other, prompted a close examination of all of the *DQ*-like cosmids with a view to piecing together a restriction map of the *DQ* region of the sheep MHC.

Cosmids 4, 39.1 and 22.1 clearly overlapped as shown in Figure 3.1. This cluster brought together the *DQA1* gene on cosmid 4, the *DQB* gene shared by all three

clones and the *DQA2* gene on cosmid 22.1. The nucleotide sequences of the *A* genes in cosmids 22.1 and 2/9.2 indicated that they were both *DQA2* genes and were probably allelic. This was supported by the phylogenetic analysis. Close inspection of the restriction maps revealed shared patterns of sites between cosmids 22.1 and 2/9.2 (Figure 3.1). Interestingly, these sites were not in the immediate vicinity of the genes themselves. Nine sites shared by 22.1 and 2/9.2 are indicated by closed circles in Figure 3.1. Similarly, cosmid 62, which contained a *DQB* gene, and cosmid 2/9.2 were linked by the 13 restriction sites marked with open boxes on the right hand side of cosmid 2/9.2 and on the left hand side of cosmid 62. As it was important to be sure of this connection, maps of cosmids 2 and 62 for the relatively infrequently cutting restriction enzymes Xhol, Hpal, Smal and Kpnl, were produced to confirm the overlap between these clones. The sites for these enzymes matched perfectly (Figure 3.10). The *B* gene in cosmid 62 has therefore been designated *DQB2* because of its proximity to the *DQA2* gene.

In this way, the *DQA1* gene on cosmid 4 was linked through *DQB1* and *DQA2* to the *DQB2* gene in cosmid 62. The distance between the 5' end of the *DQA1* gene and the 5' end of the *DQB2* gene was approximately 75 kbp. A schematic map of the *DQ* sub-region which includes cosmid 309 and the cosmids containing the expressed *DQ1* genes is shown is shown in Figure 3.11, and extends this linear tract of DNA to 130 kbp.

3.7 The transcriptional orientation of the sheep *DQ* genes.

3.7.1 The *A* genes.

To determine the transcriptional orientation of the *DQA* genes, cosmids 4, 39.1, 22.1, 9.2, 2 and 62 were digested with Hindlll (Figure 3.12) and with Sac1 (Figure 3.13) and probed with the 600 bp Pvull/Sac1 fragment which contained exon 2 from

Figure 3.10

Restriction maps of cosmids 2 and 62 from sheep 2, for the relatively infrequently cutting enzymes Xho1 (X), Kpn1 (K), Sma1 (S) and Hpa1 (H). The positions of the *DQA2* and *DQB2* genes are given by the open and filled boxes respectively.

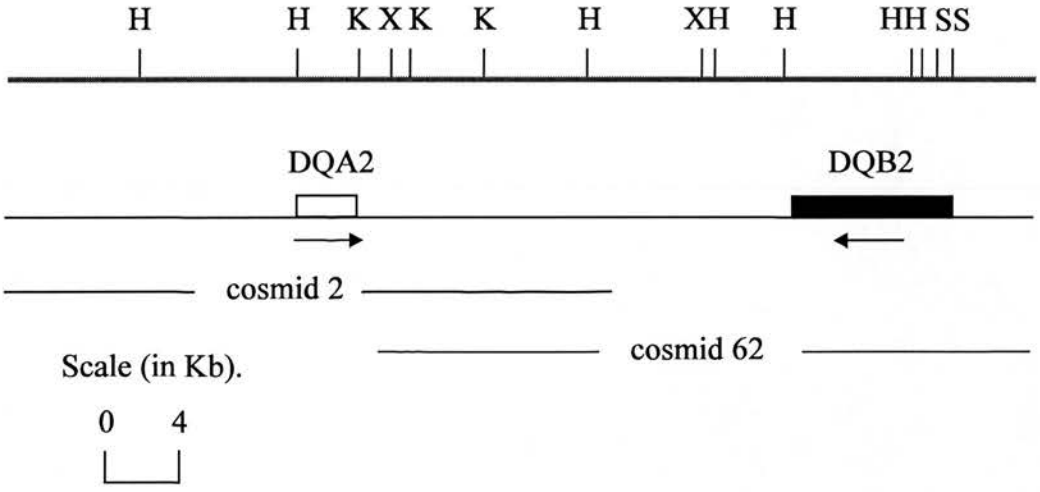


Figure 3.11

A schematic map of the sheep DQ sub-region based on overlapping cosmid clones and sequence data. See text for details.

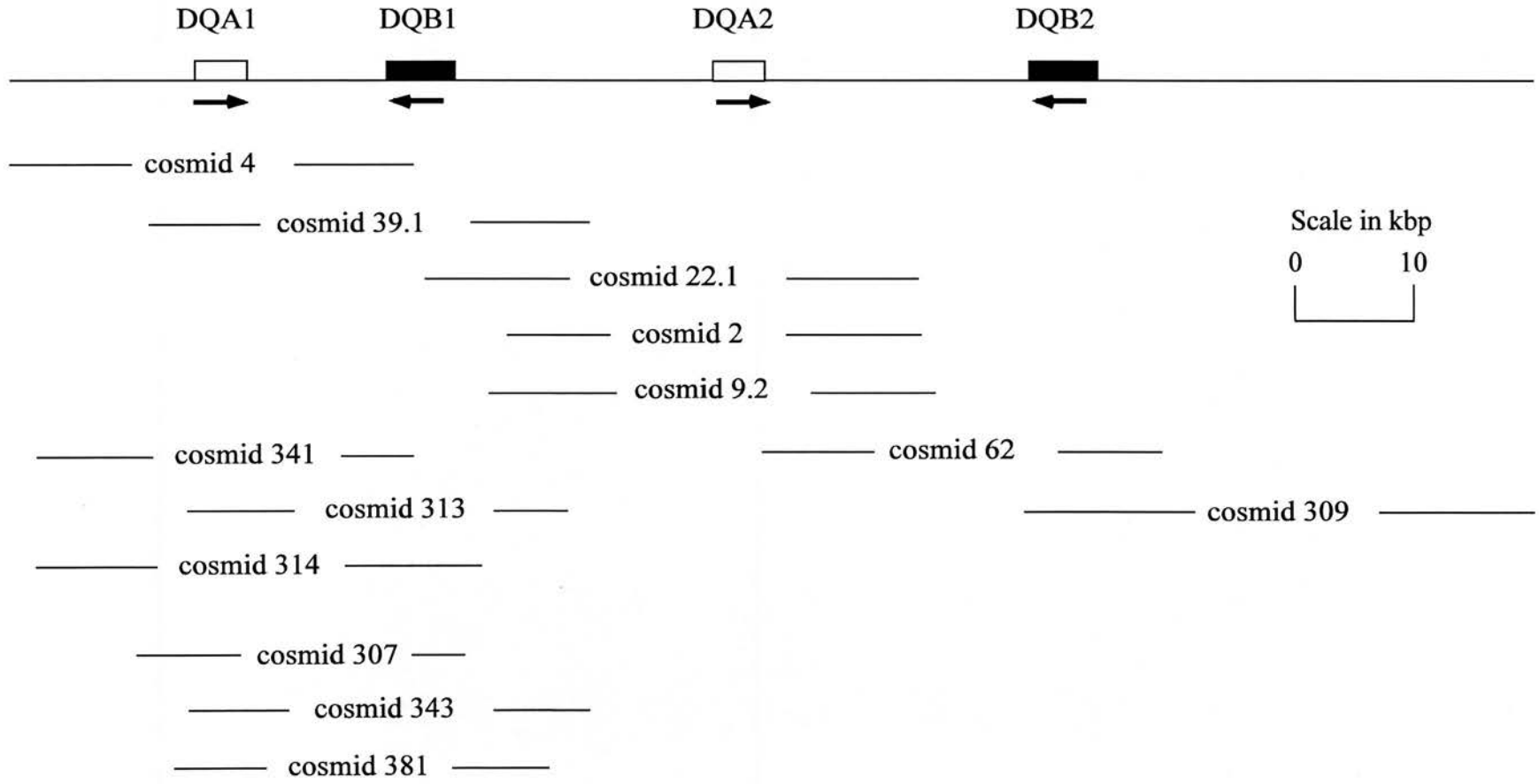
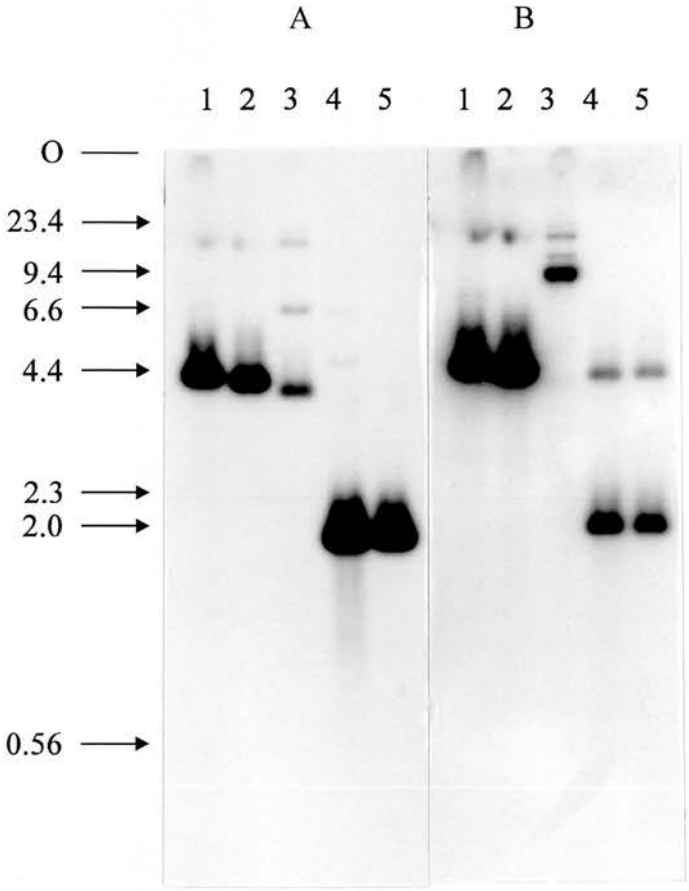


Figure 3.12

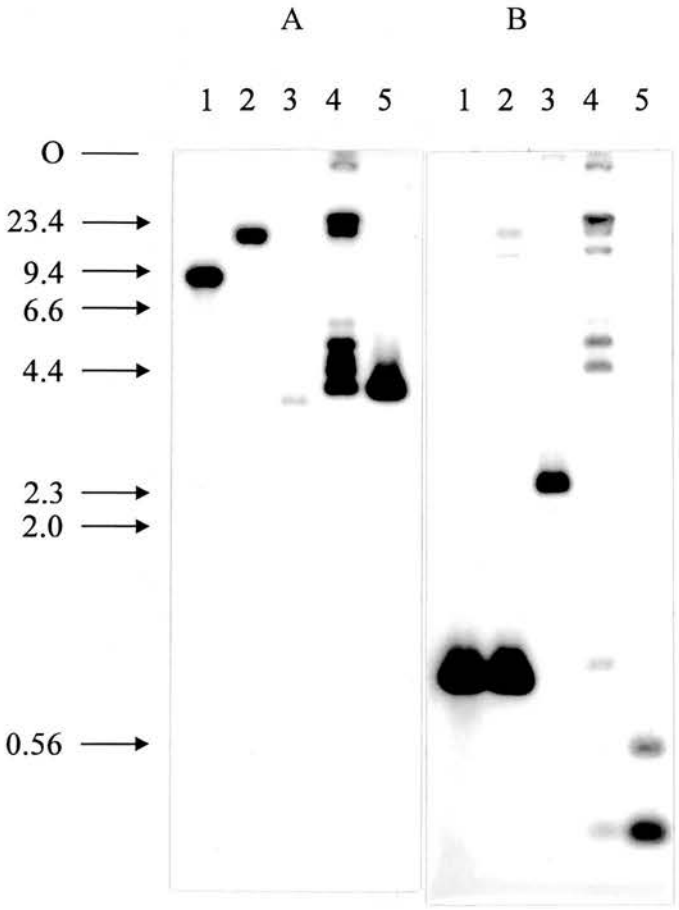
Blots of HindIII restriction digests of the cosmids from sheep 1 and sheep 2 which contain *DQA* and *DQB* genes. The blots were hybridised to A, a *DQA1* exon 2 probe and B, a *DQA1* exon 3 probe. Final washing was in 0.2xSSC, 0.1%SDS at 55^0C. The positions of the λ HindIII molecular weight markers (kbp) are indicated by arrows.



Track No.	Cosmid	Track No	Cosmid
1	4	4	9.2
2	39.1	5	2
3	22.1		

Figure 3.13

Blots of SacI restriction digests of the cosmids from sheep 1 and sheep 2 which contain *DQA* and *DQB* genes. The blots were hybridised to A, a *DQA1* exon 2 probe and B, a *DQA1* exon 3 probe. Final washing was in 0.2xSSC, 0.1%SDS at 55^0C. The positions of the λ HindIII molecular weight markers (kbp) are indicated by arrows.

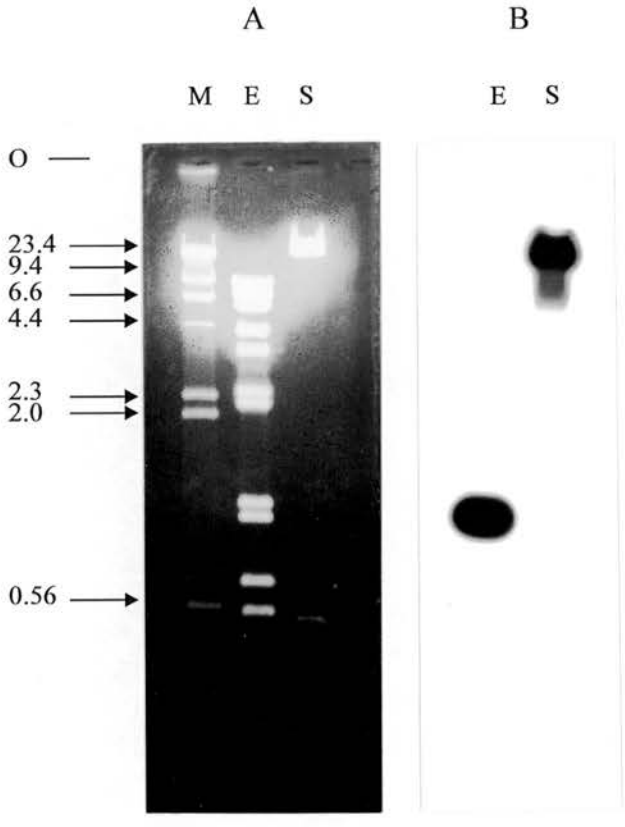


Track No.	Cosmid	Track No	Cosmid
1	4	4	9.2
2	39.1	5	2
3	22.1		

Figure 3.14

A. Ethidium bromide stained gel of cosmid 62 digested with EcoRl (E) and Smal (S). The positions of the λ Hindll markers (M) are indicated by arrows.

B. Southern blot of the gel hybridised to a *DQB* exon 3-specific probe. The blot was washed at 65^0C in 0.2 x SSC, 0.1% SDS and autoradiographed for 30 minutes at room temperature



cosmid 4, and with the 400 bp Sac1 fragment which contained exon 3 from cosmid 2 (Figure 3.3). In the case of the *DQA1* gene in cosmids 4 and 39.1 digested with Sac1 (Figure 3.13), the exon 3 probe hybridised to a shared 1 kbp fragment which could only be that in the middle of cosmid 4. The exon 2 probe hybridised to the 9.5 kbp fragment in cosmid 4 which maps to the left of the 1 kbp fragment. Therefore the direction of transcription of the *DQA1* genes is left to right as drawn.

As shown in Figure 3.12, both of the probes hybridised to a shared 2 kbp HindIII fragment in cosmids 2 and 9.2. Only the exon 3 probe hybridised to a shared 4 kbp fragment which maps to the right of the smaller fragment, giving the transcriptional direction. This was confirmed by the Sac1 digests in Figure 3.13, where the exon 2 probe hybridised to the 4 kbp Sac1 fragment in the middle of the clone, while the exon 3 probe hybridised mainly to the small Sac1 fragments to the right of the 4 kbp fragment.

3.7.2 The *B* genes.

As can be seen in Figure 3.1, the *DQB1* gene was split by the cloning procedure between cosmids 4 and 22.1. However, the two SmaI sites which define a 500 bp fragment containing exon 2, are present only in cosmid 22.1. Cosmid 4 does however hybridise to an exon 3 specific probe and hence the direction of transcription of *DQB1* is right to left as drawn, ie the *DQA1* and *DQB1* genes are orientated tail to tail.

In the case of the *DQB2* gene in cosmid 62, the second exon is once again flanked by two SmaI sites. However, as shown in Figure 3.14, an exon 3 specific probe hybridised to a 1.2 kbp EcoRI fragment which could only be one of those to the left of the SmaI fragment, there being no EcoRI sites in the pTL6 vector. Hence transcription proceeds right to left as drawn, ie. the *DQA2* and *DQB2* genes are also in a tail to tail orientation.

3.8 Expression of sheep *DQ* genes

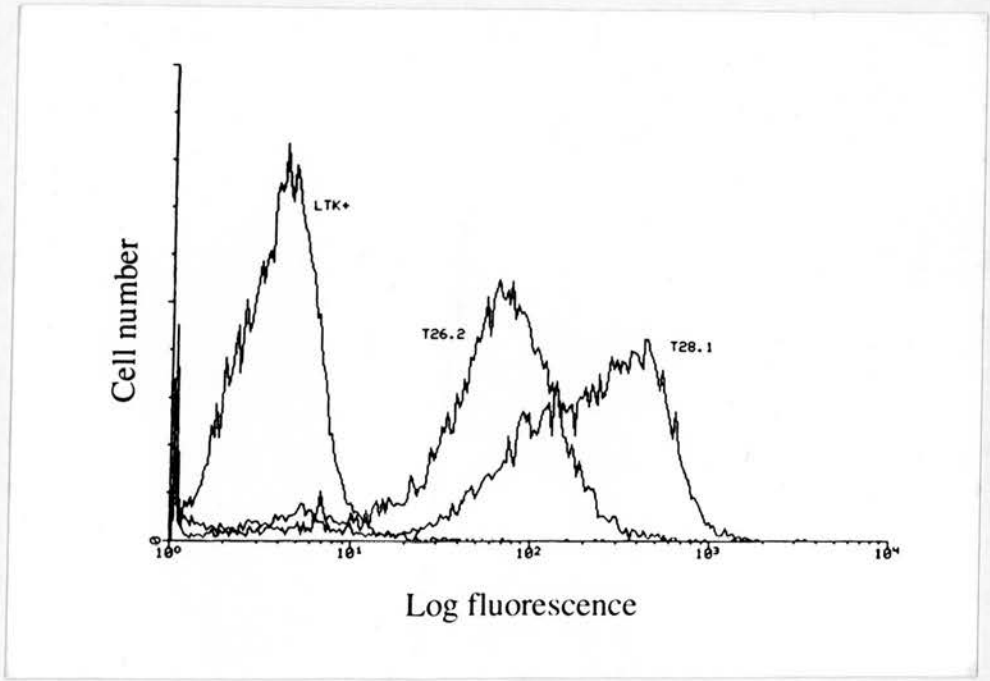
3.8.1 *DQ1* genes

Each of the cosmids from the pCos8 library which contained an *A/B* gene pair (Figure 3.3), expressed a class ll molecule at the surface of the mouse L cell following transfection of the cosmid DNA. Expression was detected using the anti-sheep class ll monoclonal antibodies as described in section 2.4.2. Five of the six cosmids in Figure 3.2 expressed high levels of class ll following only one round of cell sorting on FACS. Stable cell lines DQ-T28.1 and DQ-26.2, were established which expressed each of the two alleles, (Figure 3.15 taken from Wright et al. 1994). Ballingall et al. (1995), showed that these molecules were recognised by a different sub-set of the available monoclonal antibodies than were molecules of the *DR* isotype which had previously been established as a stable L cell line DR-T8.1 (Ballingall et al. 1992). Moreover, some mabs were able to distinguish between the two allelic forms. The results are summarised in Table 3.2 where the result of testing each mab for its reactivity with the transfected L cell is assessed on a +/- basis. Clearly mabs SBUII 38.27 and VPM1, 4, 16, 36, 40, 41, 44 and 45 were *DQ* specific. However, VPM 40, 41, 44 and 45 only reacted with one allelic form, DQ-T26.2. The sixth clone, 343, was expressed but at a much lower level. A cell line was produced, DQ-T37.3, which reacted with DQ-specific monoclonal antibodies but with an altered specificity (Table 3.2). Inspection of the restriction map of this clone, Figure 3.3, indicated that the *A* gene was near the end of the clone and could possibly be truncated by the vector. To determine whether or not this was the case, a sequencing primer (17mer) was constructed which was complementary to the sequence of the right arm of the vector pCos8, adjacent to the cloning site about 40 bp into the vector. The right hand end of the cosmid insert was sequenced by denaturing 16 μg of cosmid DNA, and annealing to this primer. This demonstrated that the *A* gene in cosmid 43 was in fact truncated 46 bp upstream of the second exon and was therefore missing exon 1 and all of the 5' regulatory elements, promotors etc. The alpha chain of the protein therefore lacked

Figure 3.15

FACS analysis of mouse L cell lines Ltk+, T26.2 and T28.1. Sheep class ll expression

was detected with a cocktail of SBUll mabs (28.1, 37.68, 38.27, 42.40 and 49.1) followed by FITC-IgG antimouse Ig as a second-stage reagent. Ltk+ was a negative control transfected with the thymidine kinase gene only. T26.2 was transfected with cosmid 314 from cluster 1 and T28.1 was transfected with cosmid 381 from cluster 2 (Figure 3.3).



part of the membrane distal domain and the leader sequence. Quite how this gene expressed a protein which was able to be transported to the cell surface has not been established. Interestingly, the truncation abrogated reactivity with some of the mabs which reacted with the intact protein expressed from the complete clones (Table 3.2). The *DQ1* genes from sheep 1 and 2 failed to express in the mouse L cell system.

Table 3.2 The specificity of the panel of monoclonal antibodies for the *DQ* and *DR* expressing L-cell lines (adapted from Ballingall et al. 1995).

Monoclonal antibody	Transfectant DQ-T28.1	Transfectant DQ-T26.2	Transfectant DR-T8.1
SW73.2	+++	+++	+++
SBUII 28.1	+++	+++ (+)	-
SBUII 42.20	-	+++ (-)	+++
SBUII 49.1	+++	-	+++
VPM 38	++	++ (-)	+++
VPM 46	++	++ (-)	+++
SBUII 38.27	+++	+++	-
VPM 1	++	+++	-
VPM 4	+++	+++	-
VPM 16	+++	+++ (-)	-
VPM 36	+++	+++	-
VPM 40	-	+++	-
VPM 41	-	+++	-
VPM 44	-	+++	-
VPM 45	-	+++ (-)	-
VPM 47	-	-	+++
SBUII 37.68	-	-	+++
VPM 54	-	-	+++
VPM 57	-	-	+++
VPM 58	-	-	++
VPM 59	-	-	++
VPM 37	-	-	-
VPM 43	-	-	-

+++ Strong, ++ variable, + weak binding, - no binding of individual monoclonal antibodies to the transfected L-cell line relative to the tk^+ L-cell control. Differences in the binding of the mAbs to the truncated *DQ* L-cell line DQ-T37.3, when compared to the complete transfectant DQ-T26.2, are indicated in parenthesis.

3.8.2 *DQ2* genes.

There was no expression of the *DQ2* genes following transfection (Ballingall, personal communication). The isolation of *DQA2* cDNA clones from sheep mRNA provided excellent evidence for their transcription (Fabb et al. 1993). Indirect evidence was obtained for the expression of *Ovar-DQ2* genes *in vivo*. Blood samples were obtained from 50 merino crosses and genomic DNA was prepared. 15 μg of each DNA sample were digested with EcoRl. Following electrophoresis, on an agarose gel, they were blotted to nylon membranes and hybridised to ^{32}P-labelled probes as described in sections 2.3.11, 2.3.13.2, 2.13.14 and 2.13.15. The probes used were the 600 bp Sac1/Pvull fragment containing the second exon from the *DQA1* gene in cosmid 4, and the equivalent fragment from the *DQA2* gene in cosmid 2. The blots were washed at high stringency. The analysis was repeated for one *DQA1* positive and two *DQA1* negative animals. The resulting autoradiographs are shown in Figure 3.16. Clearly two of these sheep were *DQA1* negative, while all three were *DQA2* positive. Peripheral blood monocytes were prepared from these sheep, labelled with the *DQ*-specific mab VPM36 (Dutia et al. 1990), and analysed by FACScan. The result is shown in Figure 3.17. Interestingly, there was a population of positive cells in the sheep which were genotypically *DQA1* negative. This could be interpreted as evidence for expression of the *DQ2* genes.

To prove the identity of the antigen which was being recognised by mab VPM36, a blood leucocyte lysate was prepared by Dr Bernadette Dutia, R(D)SVS, Edinburgh and passed over an affinity column containing immobilised mab VPM36. The bound antigen was eluted from the column and sent off for N-terminal amino acid analysis. Unfortunately, the N-terminus of the antigen was found to be blocked and so no sequence was obtained. This experiment has yet to be repeated.

Figure 3.16

Genomic Southern blots of EcoRl digests of DNA from three Merino x Suffolk sheep, 1, 2, and 3.

A. Autoradiograph of a blot hybridised to a sheep *DQA1* probe (the 500 bp Pvull/Sacl fragment containing exon 2 from cosmid 4 (Figure 3.5).

B. Autoradiograph of the same blot hybridised to a *DQA2* probe (the 500 bp Pvull/Sacl fragment containing exon 2 from cosmid 2 (Figure 3.5).

Membranes were washed at 65^0C in 0.2 x SSC, 0.1% SDS and autoradiographed for three nights at -70^0C. Arrows indicate the positions of the smaller λ Hindlll molecular weight markers (kbp).

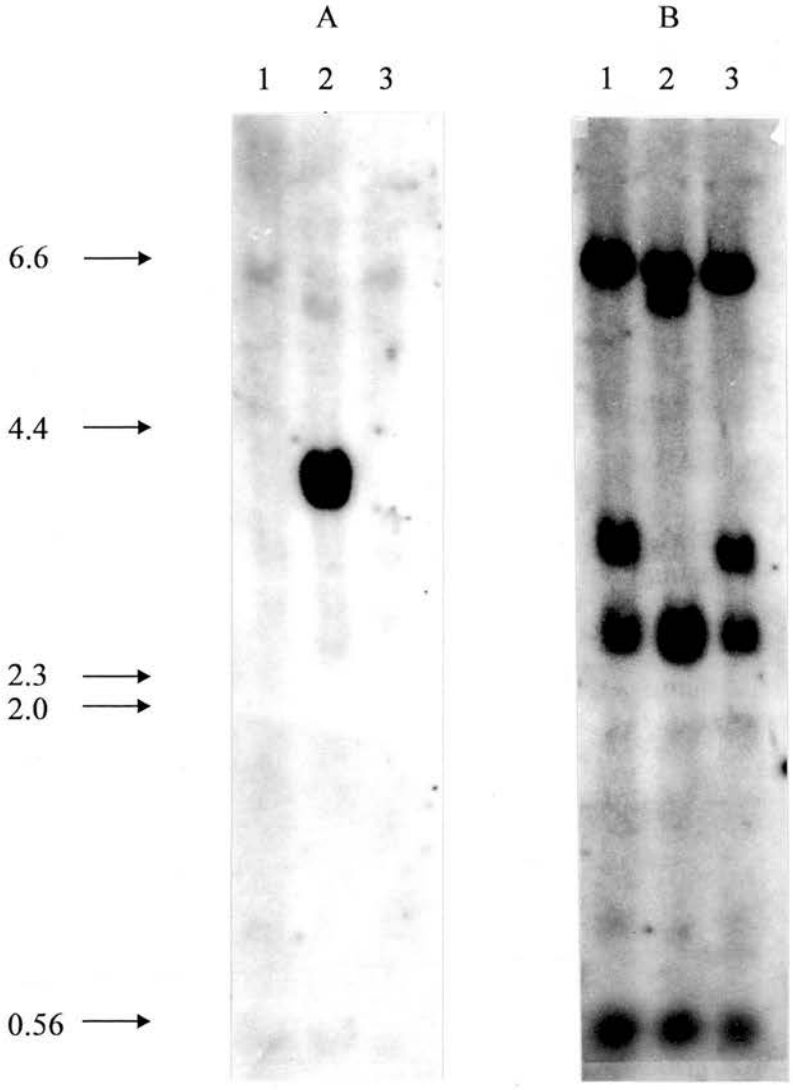
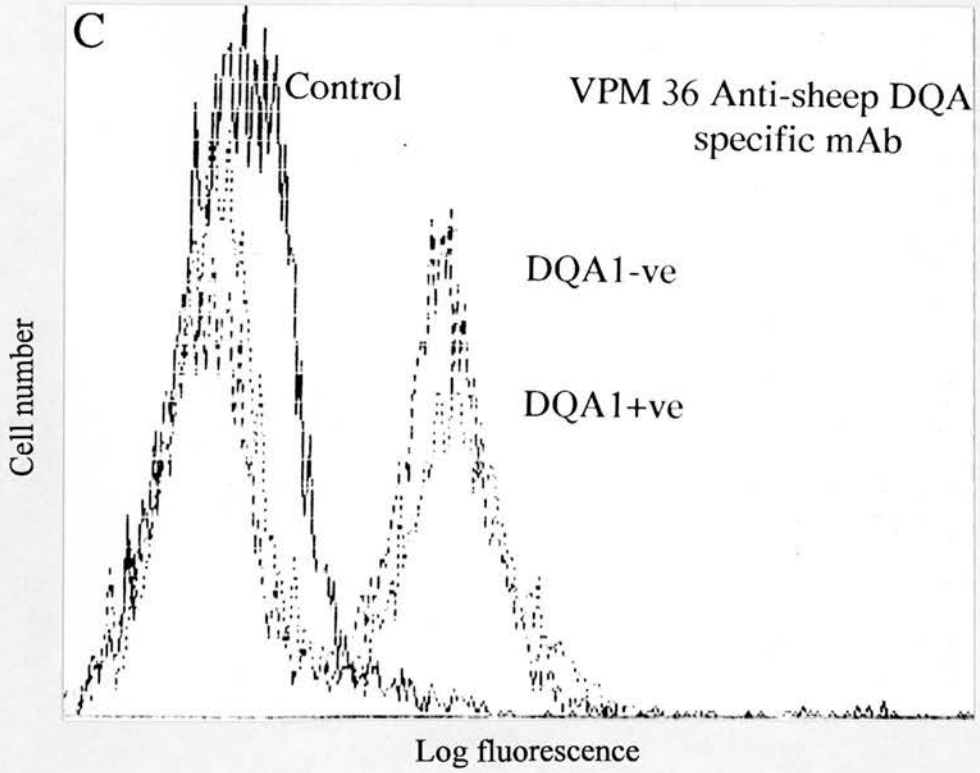


Figure 3.17

FACS analysis of peripheral blood mononuclear cells obtained from two of the merino x suffolk sheep in Figure 3.16, the *DQA1* negative sheep 1, and the *DQA1* positive sheep 2. The control population of cells were labelled with the second stage FITC-conjugated anti-mouse Ig only. Sheep 1 and 2 were labelled with mab VPM 36, a *DQA* specific monoclonal antibody (Dutia et al. 1993), followed by FITC-conjugated anti-mouse Ig.



3.9 Discussion

In this study we have identified and expressed two alleles at the *Ovar-DQ* locus equivalent to *HLA-DQ1*. The sequence homology between these *Ovar-DQ1* genes, the genes in cosmids 4 and 39.1, and the sequence designated *DQA1* by Scott et al. (1991a) suggest that they are all alleles at this locus. Two divergent *DQA* loci have been identified and sequenced in sheep (Scott et al. 1991a, Wright et al.1994). In cattle, evidence for a second *DQA* locus comes from RFLP studies (Sigurdardottir et al. 1991b). The cluster of three overlapping cosmids 4, 39.1 and 22.1 contained two linked *DQA* genes separated by a *B* gene. It was initially thought that the *A* gene in cosmid 22.1 corresponded to the *DQA3* described in some *Bota* class II haplotypes (Andersson 1988). However, the sequence data shown here suggests that the *A* gene in cosmids 2/9.2, 22.1, and the gene designated *DQA2* by Scott et al. (1991a), are alleles at a second *DQA* locus *DQA2*. By comparison of the restriction maps of cosmids containing *Ovar-DQ* genes, short regions of identity were used to link the various clones together. It must be emphasised that this map has been constructed from cosmids isolated from two unrelated sheep which were both heterozygous at the various loci. The map presented here is similar to that of the *HLA-DQ* sub-region (Campbell and Trowsdale 1993), but is more extensive than that of the mouse *I-A* region in which the *A* and *B* genes have not been duplicated. The distance between the expressed *Ovar-DQA1* and *B1* genes, at 11 kbp, was similar to that separating the corresponding *HLA-DQ1* (Kappes and Strominger 1988), and *Bota-DQ* (Groenen et al. 1990) genes. However, the physical separation between the *Ovar-DQA2* and *B2* genes, at 25 kbp, was much greater than that between the corresponding HLA genes, the only other pairing available for comparison. This large separation between *Ovar-DQA2* and *B2* genes explains why cosmids containing both these genes were not identified. Overall, the *Ovar-DQ* sub-region is more compact than the *HLA-DQ* sub-region. The distance between *Ovar-DQB1* and *DQA2* at 22 kbp, is considerably less than the 70 kbp separating the equivalent *HLA-DQ* genes (Campbell and Trowsdale 1993). No ovine equivalent of the *Bota-DQA3* (Andersson 1988) nor the *HLA-DQB3* pseudogene which is an isolated third exon (Ando et al. 1989) has been identified.

During the course of this study, a number sequences of the second exons of *DQB* genes have been reported from sheep (Scott et al. 1991b, Fabb et al. 1993, Van Oorschot et al. 1994) and from cattle (Groenen et al. 1990, Xu et al. 1991, Sigurdardottir et al. 1992b, Xu et al. 1994, Dikiniene and Aida 1995). Van Oorschot et al. (1994) described 10 new sequences from fragments amplified by PCR from sheep genomic DNA, thus further demonstrating the high degree of polymorphism in the sheep *DQ* region. In some animals, the PCR primers amplified three *DQB* sequences, confirming the presence of at least two *DQB* loci. However, it was difficult to assign the sequences to either the *DBQ1* or *DQB2* loci.

In cattle, the number of *DQA* and *DQB* genes varies between the various haplotypes (Andersson and Rask, 1988, Sigurdardottir et al. 1992b). Haplotypes *DQ*1, 2, 3, and 4 contain a single *DQA* and a single *DQB* gene, 5-12 contain two *DQA* and two *DQB* genes and haplotype 13 contains two *DQA* genes and a *DQB* gene. Sigurdardottir et al. (1992b) described 16 *DQB* sequences from 14 animals which contained each of these haplotypes. The *B* gene in the non-duplicated haplotype they designated *DQB1*, the second *B* gene in the duplicated haplotype was then called *DQB2.* However, the sequences of the *DQB2* genes in haplotypes 9-12 were very similar to *DQB1*. The *DQB1* and *DQB2* genes in haplotype 9A were almost identical to the clones Q1 and Y1, respectively, previously described by Groenen et al. (1990). The *B* gene, Q1 was obtained on a bacteriophage clone and was situated 11 kbp from a *DQA* gene which is 92% similar to the *DQA1* genes described here. There is therefore an interesting sheep/cattle difference since it is the *DQ2* genes which are present in sheep which contain a non-duplicated gene pair (Scott et al. 1991a, Fabb et al. 1993, present study). Furthermore, the MGQ amino acid motif at position 12-14 which appears to be characteristic of *Bota-DQB2* alleles (Sigurdardottir et al. 1992b, Xu et al. 1994), occurs at the same position in one of the expressed sheep *DQB1* alleles (Figure 3.9).

The duplication event which led to the presence of two *DQ* loci in man has been estimated to have occurred approximately 20 million years ago (Trowsdale et al. 1985), and therefore independently of that in sheep and cattle who last shared a

common ancestor with man approximately 75 million years ago (Trowsdale et al. 1985, Shoshani 1986). The sheep and cattle sequences are more closely related to each other than either is to the human sequence as would be expected given that sheep and cattle last shared a common ancestor approximately 20 million years ago (Irwin et al. 1991). In some cases the sheep sequences are more closely related to cattle sequences than to other sheep sequences (Van Oorschot et al. 1994 and Figures A1.1 and A1.2), reflecting the trans-species mode of evolution of MHC genes in which ancestral alleles have been passed on to the new species (Klein 1987).

The generation of stable L cell lines expressing different alleles of *DQ1* genes has been useful in the further definition of the specificities of two panels of anti-sheep class ll mabs (Ballingall et al. 1995), the SBU series (Puri et al. 1985, 1987a,b, Puri and Brandon 1987) and the VPM series (Hopkins et al. 1986, Dutia et al. 1990, 1993). These mabs allow the clear analysis of class ll molecules on immune cell populations *in vivo* (Dutia et al. 1993, Hopkins et al.1993).

The expression of the *DQ1* genes in clusters 1 and 2 from the pCos8 library contrasts with the consistent failure to obtain expression of the *DQ1* genes in cosmids 4 and 39.1. This may simply have been due to truncation of the *B* gene or its associated regulatory elements. The pattern of transcription of these genes following transient transfection (unpublished data) supports this conclusion.

Use of an ovine *DQA2*-specific probe confirmed the results of Scott et al. (1991a) in identifying high levels of transcription of *DQA2* genes in peripheral blood mononuclear cells and in macrophages (data not shown). *DQB2* mRNA could also be detected following transient transfection of mouse L-cells (Ballingall, personal communication). The production of *Ovar-DQ2* mRNA both *in vitro* and *in vivo*, suggests that a protein product may be produced. Failure to detect such a product may be due to the lack of a suitable monoclonal antibody. Certainly, "pan-*HLA-DQ*" mabs were unable to react with all *HLA-DQ* expressing L-cell transfectants (Monos et al. 1995). Sequential immunoprecipitation and N-terminal sequence analysis

undertaken by Puri et al (1987) identified *Ovar-DQ1* molecules. However, the products of the *DQ2* locus were not identified and remain to be demonstrated. *HLA-DQ* determinants are the products of *HLA-DQA1/B1* genes (Auffray et al 1987). *HLA-DQA2/B2* genes are not transcribed at high levels in normal B cells, IFN-γ treated vascular endothelial cells or transfected L-cells (Collins et al 1984, Auffray et al 1987). Transcription of the *HLA-DQB2* gene has been shown to be silenced by three point mutations in the X and W boxes within the regulatory region immediately 5' of the initiation codon (Shewey et al. 1992).

Approximately 10% of sheep from the Merino and Romney breeds have been shown to be genotypically *DQA1* -ve (Scott et al. 1991a, Fabb et al. 1993, and the present study). Furthermore, Van Oorschot et al. (1994) suggested that those sheep which were *DQA1* negative were also *DQB1* negative. Therefore, any cell surface expression of *DQ* molecules in these sheep must be the product of a functional *Ovar-DQ2* locus. In those sheep which are *DQ1* negative, the presence of a population of peripheral blood monocytes which react with a *DQA*-specific mab is therefore intriguing and suggests that the *DQ2* locus is fully expressed in the sheep.

In cattle, transcription of both *DQB* genes from duplicated haplotypes was demonstrated using RT-PCR followed by RFLP of the amplified fragments (Xu et al. 1994). Although cell surface expression of the products of both *DQ* loci was not proven, the complexity of isoelectric focussing patterns obtained with *DQB*-specific mabs (Bissumbhar et al. 1992), suggests that both loci are fully expressed in cattle.

By combining the data of Scott et al, (1987, 1991a,b), Deverson et al, (1991), and Fabb et al. (1993) with that presented here, the *DQ* sub-region of the ovine MHC has been shown to include an expressed *DQA1/B1* locus with *A* and *B* genes separated by approximately 11 kb and orientated in a tail to tail manner. The *DQ* genes are highly polymorphic. We have shown that the *DQ1* locus is linked to a *DQ2* locus with 22 kbp separating *DQB1* from *DQA2*. A similar distance separated *DQB2* from *DQA2*.

The *DQ2* genes are transcribed but cell surface products remain to be formally demonstrated.

EMBL nucleotide sequence database accession numbers for the sequences presented here are; *DQA1* gene in cluster 2, Z28418; *DQA2* gene in cosmid 2, Z28419; *DQA1* gene in cosmid 4, Z28420; *DQA2* gene in cosmid 22.1, Z28421; *DQB1* gene in cluster 1, Z28422; *DQB1* gene in cluster 2, Z28423; *DQB1* gene in cosmid 39.1, Z28424; *DQB2* gene in cosmid 62, Z28425; *DQA1* gene in cosmid 39.1, Z28518; and *DQB2* gene in cosmid 09, Z28523.

Chapter 4

The DR sub-region of the sheep MHC

4.1 Introduction

The *DR* sub-region of the human MHC is positioned at the telomeric end of the class ll region and extends over approximately 200 kbp (Campbell and Trowsdale 1993). The *DR* sub-region contains a single and functional *A* gene which is effectively monomorphic and codes for the α chain of the *DR* molecule at the cell surface (Lee et al. 1982a). The *DR* sub-region is unique in showing haplotype-specific variation in the number of *B* genes (WHO Nomenclature Committee 1991). There are five major haplotypes each of which contains an expressed *DRB1* gene together with up to four other *B* genes. In any one haplotype, there are up to two expressed *B* genes and varying numbers of pseudogenes. The *DRB1* locus is highly polymorphic with over 100 alleles now recognised, while the other expressed genes are largely oligomorphic (Bodmer et al. 1994).

In the other well characterised MHC, that of the mouse, the region equivalent to *HLA-DR* has been designated *H-2I-E,* originally identified by recombinational analysis (Murphy 1980). Cosmid cloning has shown the *H-2I-E* region to be much less extensive than *HLA-DR* and it is contained within only 40 kbp (Steinmetz et al. 1982a, 1986). The *H-2I-E* sub-region contains a single, monomorphic *A* gene, *E*α and one *B* gene, $E\beta_2$ (Steinmetz et al. 1982a, Denaro et al. 1985, Braunstein and Germain, 1986). Unusually, the other *DRB*-like gene in the mouse, *E*β, maps within another sub-region, *H-2I-A,* which is the mouse equivalent of *HLA-DQ* (Steinmetz et al. 1982a, 1986). *E*β is expressed and together with *E*α codes for the αβ heterodimeric class ll protein at the cell surface.

Prior to the commencement of the work described here, two RFLP studies had shown that the sheep MHC contains *DR*-like genes (Chardon et al. 1985, Scott et al. 1987). The latter workers described two bacteriophage clones which appeared to contain two different sheep *DRB* clones, however no clones of the sheep *DRA* gene had been identified.

In cattle, RFLP studies had defined 14 *DR* haplotypes and genetic studies had shown the cattle *DQ* and *DR* sub-regions to be closely linked (Andersson et al. 1986 a, b, Andersson et al. 1988, Sigurdardottir et al. 1988). However, the only nucleotide sequences available were of two *DRB* pseudogenes (Muggli-Cockett and Stone 1988, 1989).

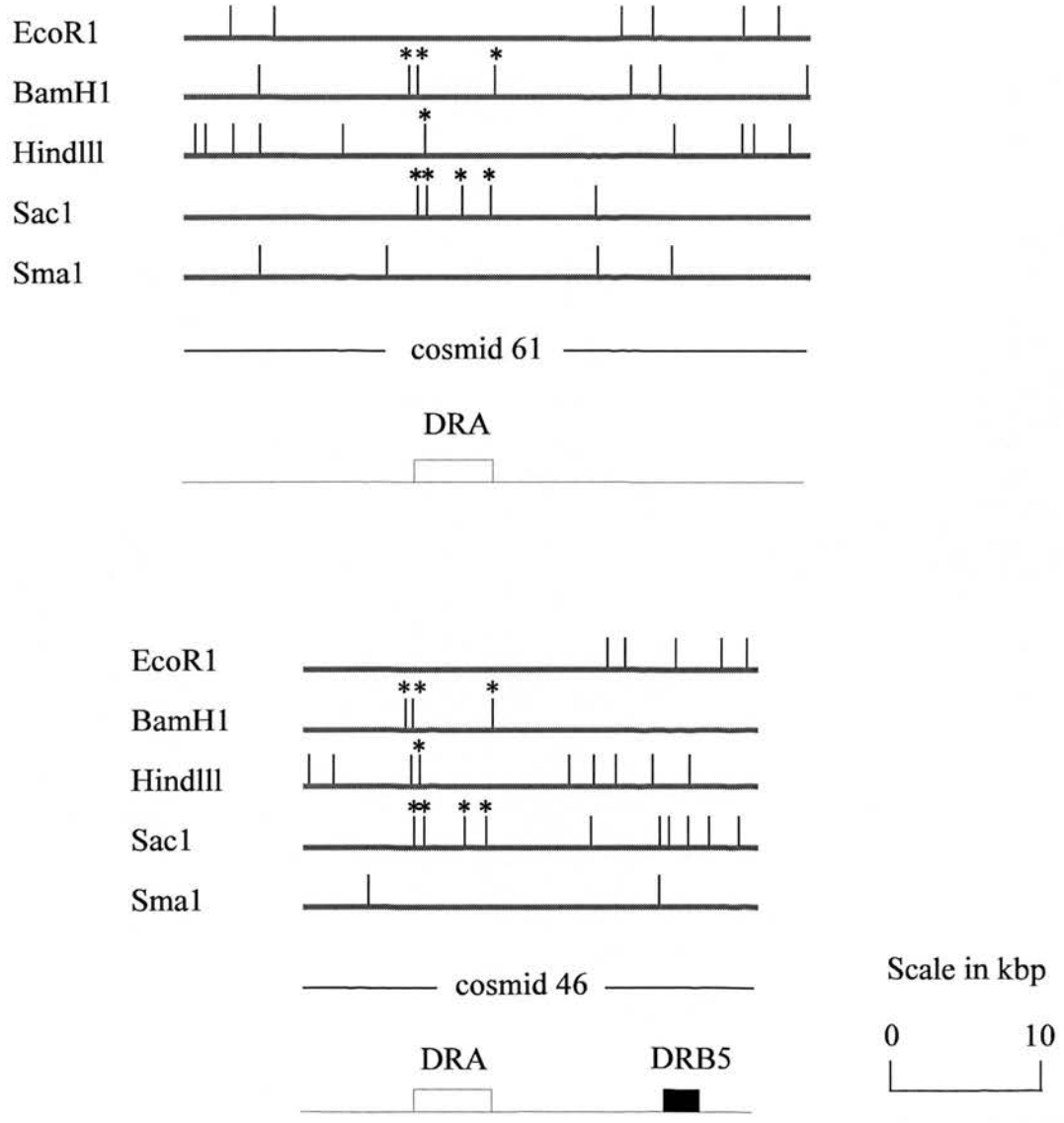
In this chapter we describe the cloning and sequencing of the sheep *DRA* gene together with a number of *DRB* genes. One of the *DRB* genes was described in a parallel study and was shown to be capable of expression when co-transfected into mouse L cells with the *DRA* gene (Ballingall et al. 1992).

4.2 The sheep orthologue of the HLA-*DRA* gene

Cosmids 46 and 61 are two clones which were obtained from the pTL6 library from sheep 2. Both clones hybridised to MHC class ll *A* gene probes, but the strongest autoradiograph signals were obtained with *HLA-DRA* as a probe. In addition, cosmid 46 hybridised to *B* gene probes. Restriction maps of the clones show that they share identical sites in the region of the *A* gene, but are markedly different outwith this region (Figure 4.1). The 5 kbp BamHl fragment which hybridised to the *A* gene probe was sub-cloned into the vector pBS+, and a set of nested deletions were made from the clone using the Exolll/S1 nuclease method as described in section 2.2.18.4. The deleted clones were sequenced from the reverse primer of pBS+ and comparison of the sequence with those in the database using the FASTA program indicated that this was the sheep orthologue of the *HLA-DRA* gene. The nucleotide sequence of the gene is given in Figure 4.2 together with the predicted amino acid translations of exons 1-4 (Ballingall et al. 1992). Subsequently, the sequence of a sheep *DRA* cDNA clone was also published (Fabb et al. 1993). The cDNA differed from the sequence reported here at only two positions. An arginine replaced a histidine at position 50 in exon 2, and a threonine replaced an isoleucine at position 109 in exon 3. A multiple alignment of the amino acid translations of the sheep, cattle (van der Poel et al. 1990,

Figure 4.1

Restriction maps for the enzymes EcoR1, BamH1, Hindlll, Sac1 and Sma1 for cosmids 61 and 46 both from the pTL6 library. Both clones contain the DRA gene as indicated by the open box, while only 46 contains the *DRB5* pseudogene indicated by the filled box. Restriction sites shared by 61 and 46 are indicated by *.



Aida et al. 1994), human (Schamboeck et al. 1983) *DRA* genes and the mouse *I-Eα* gene (Hyldig-Nielsen et al. 1983) is given in Figure 4.3. The overall structure of the gene is similar to those of mouse and man. Exon 1 encodes the signal peptide plus the first two amino acids of the membrane distal domain. Exon 2 encodes the rest of the membrane distal α1 domain. Exon 3 encodes the conserved immunoglobulin-like α2 domain, while exon 4 encodes the connecting peptide, a transmembrane domain and a short carboxy-terminal cytoplasmic domain. In sheep and cattle, the predicted signal peptide of 24 amino acids is one amino acid shorter than those of *HLA-DRA* and *H-2I-Eα* (Figure 4.3). The length of the other domains are identical, with the exception of an extra amino acid residue in the cytoplasmic domain of the mouse gene. In common with the *HLA-DRA* gene, the sheep gene is missing an amino acid at position 32 (numbered from the methionine at the start of the signal sequence), when compared with other isotypes, *DQA*, *DPA* and in the sheep, *DYA* (Chapter 6).

The percentage identities between the sheep *DRA* sequence and those of cattle, human and mouse at each of the four coding regions are given in Table 4.1. Clearly the sheep sequence is most like that of cattle with only 8 amino acid differences over the whole molecule. In exon 2, the most polymorphic region of the class ll molecule, the sheep gene shared 79, 82 and 96 % identity at the amino acid level with the mouse, human and cattle genes respectively. When the second exon of the sheep *DRA* gene was compared with those from sheep *DQA1, DQA2*, (Wright and Ballingall 1994), *DYA* (Wright et al. 1994) and *DNA* (Wright et al. 1995) genes, the corresponding figures are 51%, 52%, 54% and 42% respectively.

Only introns 2 and 3 of the sheep gene have been fully sequenced. The length of intron 2 has been well conserved between sheep, human and mouse. The sizes are 428, 489 and 486 bp respectively. The sequence identity at intron 2 between sheep and human, and sheep and mouse was 67 and 51% respectively. Intron 3 in the mouse is larger than in sheep and human, 536 versus 322 and 288 respectively. Only the sheep and human sequences were significantly similar at intron 3, sharing 68% identity.

Figure 4.2

The nucleotide sequence and predicted amino acid translation of the *Ovar-DRA* gene from cosmid 61. Coding sequences are in *capital letters* with the translation above each codon in the single letter code. Splice junctions are *underlined.* Regulatory sequences in the promoter region are *double underlined.* The putative glycosylation site is in *bold type* in exon 2, and a putative polyadenylation signal is in *bold type* at position 2315.

```
gtttttcattctttttgtttggttgtgagtttggttaattaagaagttccttttcatcaa     60
                                  S                          X1
atgcacaaagcttctcctttcctggaccctttgcaagagtctttccttagcaacagcag     120
                                 Y
tgtcatctcaaaagattttttctgattggctgcagctcaactgagtttaaatttttaatca   180
gtcagactctatgacaccctacattctcttttctcatattcctgtctgttctaatccacc    240
                                                    exon 1
                                                      M  A  I  T
ttgagctctcatcaagacacggaccctcatcaagacaccaaagaagaaaATGGCCATAAC    300
  R  V  P  I  L  G  L  F  I  T  V  L  I  S  L  Q  E  S  W  A
CAGGGTCCCAATATTAGGACTTTTCATCACTGTCCTGATCAGCCTACAGGAATCGTGGGC    360
  I  K
TATTAAAGgtaagtgctgagttcaaaactaggaataaataagttctgagagctttggaga    420
aacgaactgtggacgtttgcaagacagtct.............................     450
..................... intron 1 not sequenced.....................
cttctaccacccctgcccacacatgtgcatctcacatcctgggttcttctccatcttcct    510
                                              exon 2
                                                E  D  H  V  I  I  Q
ctcctggttcccaccctgacccccttttcttgtcgtttcagAGGACCATGTGATCATCCAG    570
 A  E  F  Y  L  N  P  E  E  S  A  E  F  M  F  D  F  D  G  D
GCTGAGTTCTATCTGAACCCTGAGGAATCAGCCGAGTTTATGTTTGACTTTGATGGTGAT    630
 E  I  F  H  V  D  M  Q  K  K  E  T  V  W  R  L  P  E  F  G
GAGATTTTCCACGTGGATATGCAGAAGAAGGAGACAGTGTGGCGGCTTCCAGAATTTGGA    690
 H  F  A  S  F  E  A  Q  G  A  L  A  N  M  A  V  M  K  A  N
CATTTTGCCAGCTTTGAGGCTCAGGGTGCCCTGGCCAATATGGCTGTGATGAAAGCCAAC    750
 L  D  I  M  I  K  R  S  N  N  T  P  N  T  N
CTGGACATCATGATAAAGCGCTCCAACAACACCCCGAACACCAATGgtacctctgctcca    810
gtcctcctagacttgggaattgtagctttaaacagatgcttgagctctttggtttttgtt    870
actgagactttcccctcaggcctattcttgtcgtatgcaaaccccacattcttatgccac    930
```

```
cggcccaaaaacttcatgaattttctcctttcttgatgtgctcacatcttgcctttgcga   990
atcatggtctctcataaagcctttgccctgagtccatgctgggggagccaggaatgaggt  1050
cattgaatatgttatgcccagtactggttccctttcatgggggagggaataagagctggg  1110
tatctgagaccacagcatagattcccaagacagtttcagctactatgtcagccttgaggg  1170
                                                      exon 3
                                                        V  P
gaggaagaggggcagggctaagcagggaaggctaatttctgcgcttgtctcccagTTCCT  1230
 P  E  V  T  L  L  P  N  K  P  V  E  L  G  E  P  N  T  L  I
CCAGAAGTGACTCTGCTCCCAAACAAGCCTGTGGAACTGGGAGAGCCCAACACACTCATC  1290
 C  F  I  D  K  F  S  P  P  V  I  S  V  T  W  L  R  N  G  I
TGCTTCATTGACAAGTTCTCGCCACCCGTGATCAGTGTCACGTGGCTTCGAAATGGCATA  1350
 P  V  T  D  G  V  S  Q  T  V  F  L  P  R  D  D  H  L  F  R
CCTGTCACTGACGGAGTGTCACAGACGGTCTTCCTGCCCAGGGATGACCACCTTTTCCGC  1410
 K  F  H  Y  L  P  F  L  P  T  T  E  D  V  Y  D  C  K  V  E
AAGTTCCACTACCTCCCCTTCCTGCCCACAACAGAGGACGTCTATGACTGCAAGGTGGAG  1470
 H  W  G  L  N  E  P  L  L  K  H  W
CACTGGGGTCTGAACGAGCCTCTTCTCAAGCACTGGGgtgagaaccgcccttcagtctcc  1530
tttacttcattgctcctcctgtgatgcatgtgtctggtccttaggaccccgaggatctgc  1590
ttcacaactgctccagttctggtttctggtcttctcccattgctctgctttcccctcttc  1650
tcgtctgtaattccctgacatcatcctgtcttcctctcatttaattggtgaaaaaaaaaa  1710
tattcttttgactaagcatcatatattttgtgctaagtgctgctgcctaccacatgcatt  1770
                                  exon 4
                                    E  Y  E  A  P  A  P  L  P  E
tcttatgtactctgacttatttttccccagAGTATGAAGCTCCAGCCCCTCTCCCAGAGA  1830
T  T  E  N  A  V  C  A  L  G  L  I  V  A  L  V  G  I  I  A
CTACAGAGAATGCGGTGTGTGCCCTGGGCCTGATTGTGGCTCTGGTGGGCATCATTGCAG  1890
G  T  I  F  I  I  K  G  V  R  K  A  N  T  V  E  G  R  G  P
GGACCATCTTCATCATCAAGGGCGTGCGCAAAGCCAACACTGTTGAAGGCCGAGGGCCTC  1950
L  *
TGTGAggcacctgcaggtgagtctgctgtggtcagaggatgacatctctggaatgattcc  2110
agaggaggaaaagagtgaggaaaaggacacacgatgcctttaaaggaaaaccattcctaa  2170
agtcatggctcctgattcatcacactggacagaatcagacattgtcatcgtctgatatcc  2230
ccaagccctacattccatacgtgtcaccagagatcatgccttcggtcttggaaactatct  2290
ccagtaccaaaatagttgtttcac attaaa atagtgagtccaatgatctgggaaagtcag  2350
cctcaaaattacaacacttcctggctttatatttcctgagataagacatcagctttgaag  2410
catttccagaatgtcacttttggagggtgtagggtaactgtatgaagtggaatctcttgg  2470
```

Table 4.1. The percentage identity at both the nucleotide (Nu), and amino acid (**AA**) levels between the sheep gene in cosmid 61, and the cattle *DRA* gene (Aida et al. 1994), human *DRA* gene (Schamboeck et al. 1983) and the mouse I-Eα gene (Hyldig-Nielsen et al. 1983). Amino acid identity is given in bold type.

	Exon 1		Exon 2		Exon 3		Exon 4	
	Nu	**AA**	Nu	**AA**	Nu	**AA**	Nu	**AA**
Bota-*DRA*	99	**99**	97	**96**	97	**96**	97	**96**
HLA-*DRA*	80	**64**	88	**86**	87	**82**	84	**80**
I-Eα	72	**40**	77	**75**	79	**79**	73	**63**

Neither the sheep nor the cattle cDNAs have a consensus polyadenylation signal AATAAA, in the 3' untranslated region. Instead they both have the AGTAAA mutation (Fabb et al. 1993, Aida et al. 1994). The genomic sequence reported here has the only common variant ATTAAA at position 2315 of the sequence (Figure 4.2). However, it has not been proved that this is the polyadenylation signal. In particular, the CAPyTG downstream element critical to the cleavage reaction which precedes polyadenylation (Manley 1988), is missing in the sequence shown. The sheep and cattle sequences also differ from those of the human and mouse in having only one potential glycosylation site. The canonical asparagine-X-(threonine or serine) sequence occurs at the conserved position 78 in the α1 domain of the sheep and cattle sequences (Figure 4.3). The human and mouse sequences have an additional site at position 118 of the α2 domain. The cysteines at positions 107 and 163 form the disulphide bonds within the α2 domain of the protein chain.

Figure 4.3

Mulptiple alignment of the *DRA* genes of the, sheep, cattle, human and mouse MHC class ll regions. Dashes indicate identity. Potential glycosylation sites are in bold type. Codon insertions/deletions are indicated by a dot. * indicates a stop codon.

```
        Signal peptide
Ovar    MAITRVPILGLFI.TVLISLQESWA
Bota    -------------.----G------
HLA     ---SG--V--F--IA--M-A-----
IE      --TIGALV-RF-FIA--M-S-K---

        α1 domain
        1       10        20        30        40        50        60        70        80
Ovar    IKEDHVIIQAEFYLNPEESAEFMFDFDGDEIFHVDMQKKETVWRLPEFGHFASFEAQGALANMAVMKANLDIMIKRSNNTPNTN
Bota    ---N----------K---------------------G-----------------------------------------------
HLA     ---E------------DQ-G---------------A--------E---R------------I--D----E--T----Y--I--
I-E     ---E-T--------L-DKRG--------------IE-S-TI---E--AK------------I--D-----V-KE------DA-

        α2 domain
            90        100       110       120       130       140       150       160       170
Ovar    VPPEVTLLPNKPVELGEPNTLICFIDKFSPPVISVTWLRNGIPVTDGVSQTVFLPRDDHLFRKFHYLPFLPTTEDVYDCKVEHWGLNEPLLKHW
Bota    ----------------------------------------K--------------N--------------------------L----------
HLA     ------V-T-S----R---V--------T---VN-------K---T---E------E--------------S-------R------D-------
I-E     -A----V-SRS--N-----I----I-------VN-------R---E---E-----------------T---S-D-F---E-D----EEPLRKTW

        CP domain       TM domain                 CY domain
       180       190         200       210         220
Ovar    EYEAPAPLPETTE  NAVCALGLIVALVGIIAGTIFII  KGVRKANTVEGR.GPL*
Bota    -------------  ----------------------  ----------R-.----
HLA     -FD--S-------  -V------T-G-----I------  --L--S-AA-R-.----
I-E     -F-EKTL----K-  -VM-----F-G----VV-I-L-M  --IK-R-VV-R-Q-A--
```

As was indicated above, the sheep *DRA* gene described here is fully expressed in the mouse L cell (Ballingall et al. 1992). DNA sequences upstream of the initiation codon are critical in the regulation of transcription of MHC class ll genes (reviewed by Benoist and Mathis, 1990, Glimcher and Kara 1992). The sequence of the promoter region of the sheep *DRA* gene is given in Figure 4.2 where the regulatory motifs are double underlined. In common with the human *DRA* gene the sheep gene does not contain a forward CCAAT motif. However, it does contain the consensus Y box, a ten base pair motif containing a reverse CCAAT sequence 141 bp upstream of the transcription start site. Twenty two bp upstream of the Y box is the X_1 box which shows only two differences from the human sequence. Further upstream is the S box of seven base pairs which is totally conserved between sheep and man. The *HLA-DRA* gene is unique in having an octamer motif downstream of the Y box at -53 (Tsang et al. 1990). Interestingly, the sheep *DRA* gene appears to lack this regulatory box.

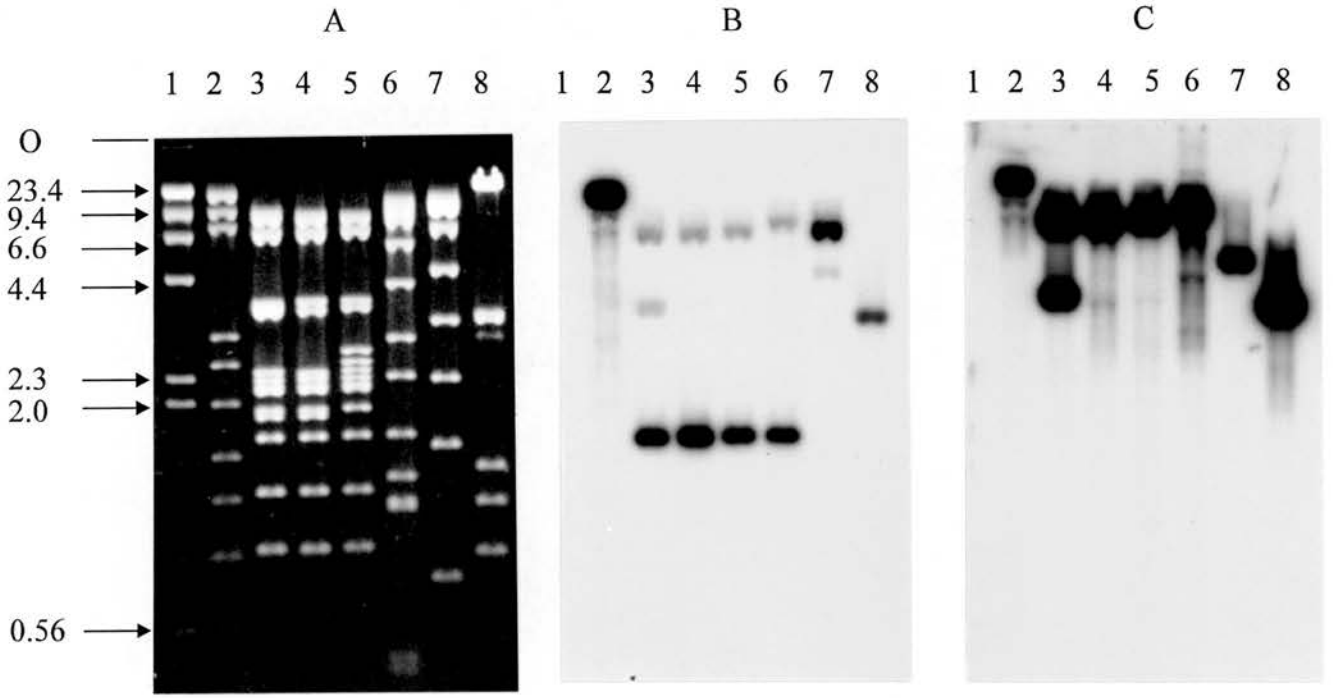
4.3 Phylogenetic relationships between the class ll *DRA* genes

The genetic relationships between the various class ll *A* genes were investigated by the construction of a phylogenetic tree using the neighbour-joining method as described in Appendix 1. The tree is shown in Figure A1.1. Clearly the sheep gene in cosmid 61 forms part of a branch which contains only the *DRA*-like genes of the human, mouse, and cattle and occurred in 2000 out of 2000 trials. Within this branch the sheep and cattle genes form a separate branch in 1866 out of 2000 trees.

4.4 Sheep MHC class ll *DRB* genes

A number of clones which hybridised in a non-specific way to human and mouse *B* gene probes were obtained from the screening of the pTL5 and pTL6 libraries. EcoRl restriction digests of those clones which have since been shown to contain *DRB*-like genes are shown in Figure 4.4. This gel was blotted to Hybond-N and hybridised

Figure 4.4



sequentially with *DRB* exon 2 and exon 3 specific probes. Photographs of the autoradiographs are also given in Figure 4.4.

4.4.1 An allele at an expressed *DRB* locus

The restriction maps of cosmid 9.5 from the pTL5 library (sheep 1) and cosmid 22 from the pTL6 library (sheep 2) were almost identical (Figure 4.5). There were only five sites unique to cosmid 22 and one to cosmid 9.5. Cosmids 9.5 and 22 hybridised to both exon 2 and exon 3 gene probes (Figure 4.4). The *B* gene in cosmid 22 was sequenced by K.T. Ballingall and was shown to be a sheep *DRB* gene (Ballingall, 1991, Ballingall et al. 1992). The second exon of the gene in cosmid 9.5 was also sequenced and was identical to that of cosmid 22. Following transfection of mouse L cells with DNA from either cosmid 22 or 9.5, in combination with the *DRA* gene described above in cosmids 61 or 46, a class ll molecule was detected at the cell surface by FACS analysis (Ballingall et al. 1992). The gene in these cosmids is an allele at an expressed *DRB* locus within the sheep *DR* sub-region and we have called this gene *Ovar-DRB1* to be consistent with the expressed *DRB1* gene in the HLA.

4.4.2 The sheep *DR* sub-region may contain up to four *DRB* pseudogenes

All of the other genes shown in Figure 4.4 failed to express in the mouse L cell system (Ballingall, personal communication). As will be shown below, in some cases this was because the genes were truncated or contained in-frame stop codons. However, the pseudogene status of all of the cloned genes has not been formally proved by sequencing. In all cases sequencing has shown the other *DRB*-like genes to be markedly different from that of the expressed locus.

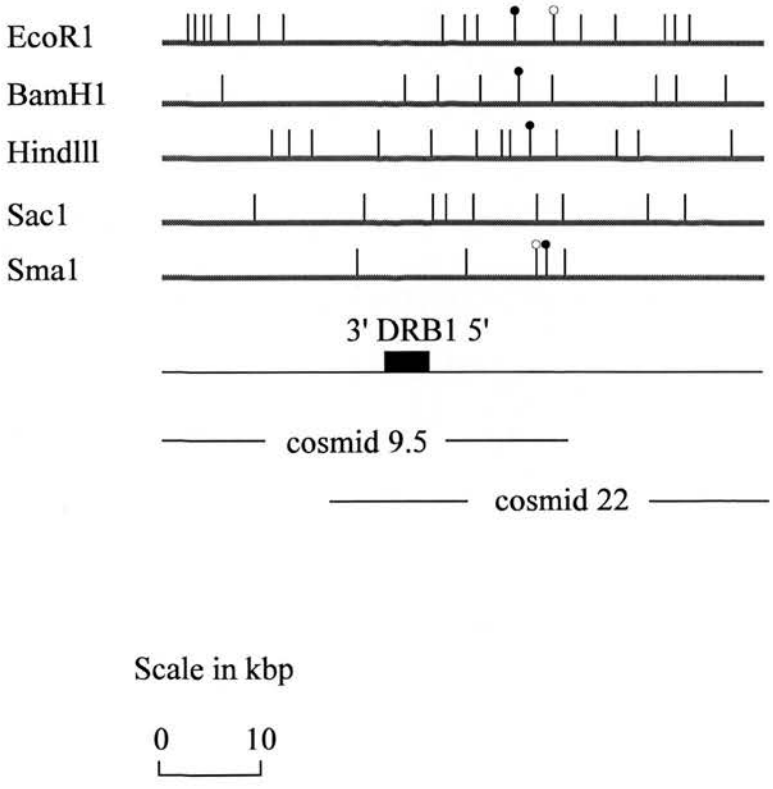
The EcoRl restriction digests of cosmids 9.1, 17.2 and 43.1 show that these clones are closely related (Figure 4.4). The clones come from sheep 1, and their restriction maps for the common mapping enzymes EcoRl, BamHl, Hindlll, Sacl and Smal reveal

Figure 4.5

Restriction maps of overlapping cosmids 22 from the pTL6 library and 9.5 from the pTL5 library for the enzymes EcoR1, BamH1, Hindlll, Sac1 and Smal.

Closed circles indicate sites unique to cosmid 22, open circles indicate those unique to cosmid 9.5. The black box indicates the position of the gene.

These clones contain alleles at an expressed *Ovar-DRB1* locus. The direction of transcription is indicated.



only three restriction site differences (Figure 4.6). The clones link together two *B* genes which for present purposes have been designated *DRB2* and *DRB3*.hybridises to both an exon 2- and an exon 3-specific probe (Figure 4.4). Cosmid 9.1 contains both genes, but the only additional hybridising fragment is exon 3-specific (Figure 4.4). Hence, *DRB3* is truncated, possibly by the cloning process, and lacks exon 2 and upstream sequences. Cosmid 32, shares an EcoRl fragment of about 1.7 kbp with cosmids 9.1, 17.2 and 43.1. This fragment hybridises to an exon 2-specific probe (Figure 4.4). Cosmids 9.1 and 32 may also share the Sacl and Smal sites in the region of *DRB2* (data not shown). The remainder of the restriction maps in the area of overlap appeared unrelated (Figure 4.6).

The 1.7 kbp EcoRl fragment from cosmid 9.1 which hybridised to the exon 2 probe was excised from an agarose gel, purified and ligated into the EcoRl site in the MCS of pBS+. DNA from the resulting clone was purified by CsCl density centrifugation and a deletion library was created using the Exolll/Sl nuclease procedure (section 2.2.18.4). Sequence data was accumulated from the reverse primer of pBS+, and 17-mer oligonucleotides were purchased to confirm the sequence on the opposite strand. The oligonucleotides around exon 2 were used to prime full length cosmid DNA (ie without sub-cloning) to show that the sequences of the genes on cosmids 17.2, 43.1 and 32, were identical to that of *DRB2*. The nucleotide sequence of *DRB2* is shown in Figure 4.7 together with its putative amino acid translation. A database search using the FASTA comparison program indicated that this was a *DRB*-like gene. To confirm this, the third exon of the gene was sequenced. Two sequencing primers (17-mers) were derived from the sequence of the third exon of the expressed sheep *DRB1* gene (Ballingall et al. 1992). The sequences of the primers and their positions are indicated in Figure 4.8. The sequence of exon 3 from *DRB2* was obtained by direct priming of cosmid 43.1 as this clone did not contain *DRB3*. Cosmid 9.1 was the only clone to contain *DRB3*. However, as 9.1 also contained *DRB2*, the Hindlll fragment of approximately 3 kbp at the left hand side of the cosmid as drawn, which contained the *DRB3* gene, was gel purified, digested with Sau3A and

Figure 4.6

Restriction maps of overlapping cosmids 9.1, 17.2 and 43.1 from the pTL5 library and cosmid 32 from the pTL6 library for the enzymes EcoR1, BamH1, Hindlll, Sac1 and Sma1.

The sites marked with a closed circle are polymorphic in clones 9.1, 17.2 and 43.1.

The EcoR1 sites marked with open circles are common to all four clones.

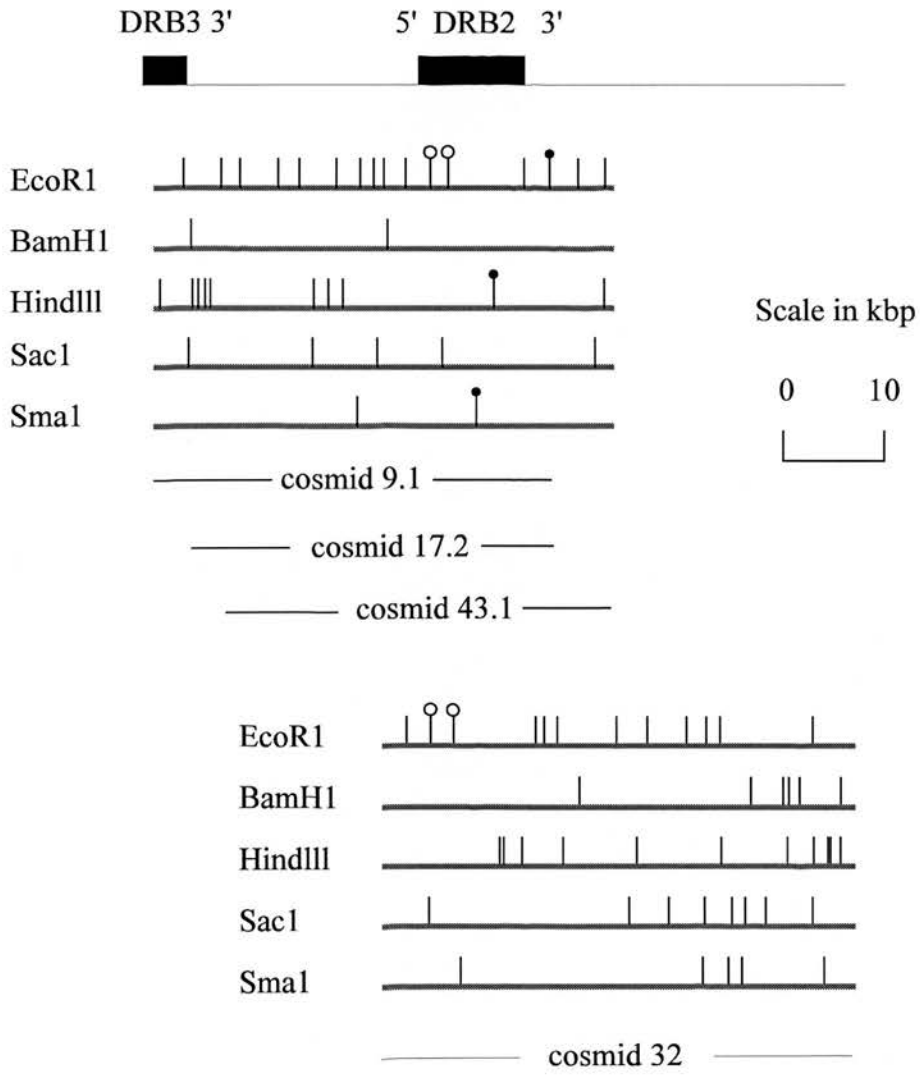


Figure 4.7

Alignment of the nucleotide sequences and predicted translations of the second exons of the *DRB* genes in cosmids 23.1 (*DRB4*) and cosmids 32/9.1/17.2/43.1 (*DRB2*). See text for details.

```
DRB4  gtcccaggaggaggcggtggtggggaccccgaccaggccacacttggcgacgccaggcga  60
DRB4  gcctggtggccaagaggagcccgatcgctcaggggcagagctcgtcttccccctgcgcgg 120
DRB4  ctggcgggggagttcggaccggggctccgaggggagcccttgggggtccccgtggccccg 180

                                                      L  F  M  V  Q
DRB4  cgcgccgggcgccaggtaaccagggcactcccgtccccgcagCACTCTTCATGGTGCAGG 240
DRB2              ccagtgatgagatctttcacttcttcacagCCCTTTTCATGCATCAGT
                                                      L  F  M  H  Q

      G  K  S  E  C  H  F  S  N  G  T  Q  Q  V  R  F  L  D  R  Y
DRB4  GCAAGTCCGAGTGTCACTTCTCCAACGGCACGCAGCAGGTGCGCTTCCTGGACAGGTACA 300
DRB2  TTAAGGGTGAGTGTCGTTTCTCCAACGGGTTGGAGCGGATGCGGTTCTTTGCCAGATACA
      F  K  G  E  C  R  F  S  N  G  L  E  R  M  R  F  F  A  R  Y

      I  Y  N  R  E  E  Q  V  R  F  S  D  L  V  G  E  Y  R  A  R
DRB4  TCTACAACCGCGAGGAGCAGGTGCGCTTCGACAGCCTCGTGGGCGAGTACCGGGCGCGGA 360
DRB2  TCTACAACACGAAGGAGGATGTGCACTTCGACAGCGATGTAGGGGAGTTCACGGCCCTGA
      I  Y  N  T  K  E  D  V  H  F  D  S  D  V  G  E  F  T  A  L

      T  E  M  G  R  P  A  A  E  R  W  N  R  W  P  D  A  L  Q  R
DRB4  CCGAGATGGGGCGGCCGGCGGCCGAGCGCTGGAACCGCTGGCCCGACGCCCTGCAGCGGG 420
DRB2  CGGAGCTGGGGAGGCCGGACGCCGAGTACTGGAACCAGCAAAAGGACTTCATGGAGCAGA
      T  E  L  G  R  P  D  A  E  Y  W  N  Q  Q  K  D  F  M  E  Q

      A  R  A  A  V  H  D  F  C  A  S  N  Y  E  F  F  A  S  R  T
DRB4  CCCGCGCCGCCGTGCACGATTTCTGCGCCAGCAACTACGAGTTCTTTGCAAGCCGCACGG 480
DRB2  TGCGGGCCAAGGTGGACACGGTGTGCAGATCCAACTATCAGGGCATTGGTAGCTTCCTGA
      M  R  A  K  V  D  T  V  C  R  S  N  Y  Q  G  I  G  S  F  L

       V  Q  R  R
DRB4  TGCAGAGGAGAGgtgagccgggatgggggtgagagatcggtggcaggggaggggagtgtg 540
DRB2  GGCAGCGCCGAGgtgagcgaggccggtgagtggcctcggggaggttgtgtgtgtgtgcgt
      R  Q  R  R

DRB4  agagcgaccgagacagagacggagagacagagagagacggagatagagagacggagagac 600
DRB2  gcgtgtgtgcaggagtgtgtgtgtgagagagagacggagagaaggacagagacagaggct

DRB4  agagagagacggagatagagagacagagacacagagacagagacacagagagaaggagat 660
DRB2  gagtccaggtgagtgctggattgtgagcaagttcaagttaggaaagttgctgtgagagga

DRB4  agagagacagagacacagagagacggagatagagggacagagacacagagacagagacac 720
DRB2  tgagactgggagtctcacatgctgtcggggagacagtgtgtgattttgtcagtgtgtgat

DRB4  acacagagacagagacacagagacagagacacacacagagacagagacacagagacacag 780
DRB2  tgtgagaggggcgaggagagcatgtggtgggggaaggggccgggggtctgagagaaccag

DRB4  agacagagacacagagtgagagacacagagagacagagagagacagagacacactgtgag 840
DRB2  aaagggtatgaggggtgggggagatgctggaaggaaagatggtagcaggctgtgggagag

DRB4  agacggagagacagagagagacggagatagagagagacacagagagacagagagagac   898
DRB2  gtgaggtggcagaaaggggaaggggatatggaggaatactgactgggatatggggttagg
DRB2  cttcctggtggctcagacattaagaatctgcctgcaatgcaggagacctgagtttgatcc
DRB2  ctggctggggaagatcctctggagaagagaatggctacccactcaagt
```

Figure 4.8

Nucleotide sequence alignment of exon 3 of the expressed *DRB* gene in cosmid 9.5 (*DRB1*), and the pseudogenes in cosmids 9.1 (*DRB2* and *DRB3*), 23.1 (*DRB4*) and 46 (*DRB5*). The primers used for sequencing are double underlined.

```
        exon 3                                                      50
DRB1    agTGGAGCCTATAGTGACTGTGTATCCTGCAAAGACCCAGCCCCTGCAGC
DRB2    ----------C-----------------------------------------
DRB3    --GT-GA----C-----------------------------------------
DRB4    ---T-------CG----------------T-----GT-G---T--ACG--
DRB5    ----------C-----------------A--------------------

                                                           100
DRB1    ACCACAACCTCCTGGTCTGCTCTGTGAATGGATTCTACCCAGGCCACATT
DRB2    ------G-------------------------T------------------
DRB3    ------G-------------------------T------------------
DRB4    ------------------------------T--------T---------
DRB5    -----G--------------------G----T-----T-TG-----TGC-

                                                           150
DRB1    GAAGTCAGGTGGTTCCGGAATGGCCACGAAGAGGAGGCTGGGGTGATCTC
DRB2    --------------------C-----T----------------------
DRB3    --------------------C-----T----------------------
DRB4    -------A--------A---------G-----------C------G----
DRB5    ---------------T----A-----G----------------------

                                                           200
DRB1    CACAGGCCTGATCCAGAATGGAGACTGGACCTTCCAGACCATGGTGATGC
DRB2    ----------------------------------------G---------
DRB3    ----------------------------T-----------G---------
DRB4    --------------CT----------------------T-----------
DRB5    ------------------C-------A----C------------------

                                                           250
DRB1    TTGAAACAGTTCCTCAGAGTGGAGAGGTCTACACCTGCCAAGTGGATCAC
DRB2    ----------------C---------------------------G---
DRB3    ------C---------C---------------------------G---
DRB4    ----T-----------G----------------------C-----G---
DRB5    ----TG-----------------------------------------G---

DRB1    CCCAGCCGGACGAGCCCTATCACAGTGGAATGGAgt
DRB2    ------------------C-----------------
DRB3    ----------GC------C-----------------
DRB4    --------C---------G------CA---------
DRB5    -----T-A----------C-----------gcagag
```

shotgun cloned into pBS+ which had been digested with BamHl. Sau3A subclones containing exon 3 were identified following colony lifts, by hybridisation to a *DRB* probe. The sequence of exon 3 from *DRB3* was then obtained using the specific primers described above (Figure 4.8).

Two further cosmid clones, 23.1 and 37.1, had been obtained from the library from sheep 1, but the *B* gene they contain had not been assigned to a specific sub-type. EcoRl restriction digests (Figure 4.4) and the restriction maps of the clones (Figure 4.9), indicated that they were not directly related to the other clones described in this chapter. However, they did hybridise to *DRB* probes (Figure 4.4). A Sacl fragment of approximately 1.4 kbp was identified which hybridised to the 300 bp Pstl fragment which contained exon 2 from the expressed *DRB1* gene in cosmid 22. The Sacl fragment was cloned in pBS+ and the fragment which contained exon 2 was sequenced using a combination of primers and has been designated *DRB4* (Figure 4.7). The sequence of exon 3 from cosmid 23.1 was obtained by directly sequencing the cosmid DNA using the primers described above and is given in Figure 4.8.

The sequence of exon 3 from the *DRB* gene adjacent to the sheep *DRA* gene in cosmid 46, was also obtained using the primers described above and is designated *DRB5* in Figure 4.8.

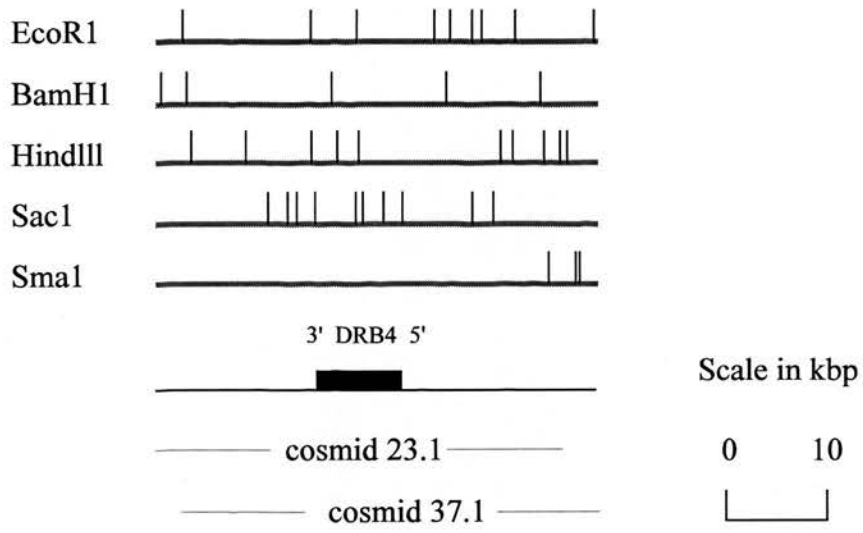
4.4.3 Sequence relationships between the *DRB* pseudogenes

The predicted amino acid sequences of exons 2 and 3 of the *DRB* pseudogenes are shown in Figure 4.10 where they are compared with those of the expressed *DRB1* gene. Overall, the *DRB1* gene shares only 60% and 55% identity at exon 2 with *DRB2* and *DRB4* respectively. *DRB2* and *DRB4* share 52% amino acid identity at the same exon. *DRB2* has lost the putative glycosylation site at position 19 of exon 2. All of the genes have retained the four conserved cysteine residues at positions 15, 79, 116 and 172 involved in intradomain sulphide bridge formation. In both of the pseudogenes the Bovidae-specific glycine at position 34 (Andersson et al. 1991b) has

Figure 4.9

Restriction maps of overlapping cosmid clones 23.1 and 37.1 for the enzymes EcoR1, BamH1, Hindlll, Sac1 and Sma1.

These clones contain a *DRB*-like pseudogene indicated by the black box.



been replaced. *DRB2* has an arginine at this position which is found in expressed red deer *DRB* genes (Swarbrick et al.1995), while *DRB4* has a threonine residue not present in any other DRB sequence. The Bovidae-specific tryptophan at position 44 (Andersson et al. 1991b) has been replaced by a valine residue in the two sheep pseudogenes which occurs in primates and red deer (Marsh and Bodmer 1993, Swarbrick et al. 1995). At position 68, the leucine characteristic of expressed *DRB* genes across species has been replaced by methionine which occurs only in expressed red deer *DRB* sequences. At exon 3, *DRB1* shares 93, 90, 80 and 84% amino acid identity with *DRB2*, *DRB3, DRB4* and *DRB5* respectively. *DRB5* has an in-frame stop codon at position 152 in place of the conserved trptophan residue and has lost the splice consensus sequence which defines the end of exon 3.

The *DRB* genes described here were included in a phylogenetic analysis of selected *B* genes. Two trees were constructed, one using the amino acid translations of exon 2 sequences (Figure A1.2), and one using exon 3 (Figure A1.3). The second exons of sheep genes *DRB2* and *DRB4* grouped together on a branch with those of the cattle genes *Bota-DRB2* (Muggli-Cockett and Stone 1989) and *Bota-DRB1* (Muggli-Cockett and Stone 1988). They were quite separate from the group containing the expressed sheep *DRB1* gene. Indeed they appear to be an early off shoot from the branch leading to the *DQ* genes. The branch containing the sheep and cattle *DRB2* genes was highly significant and occurred in 99.5% of trials. The identity at the amino acid level between these genes was 92%. The juxtaposition of the sheep *DRB4* and the *Bota-DRB1* genes on the other hand was not significant. The identity at the amino acid level between these genes was only 55%.

In the tree constructed from exon 3 sequences, all of the sheep and cattle *DRB* sequences grouped together and were distinct from the mouse *I-E*β and human *DRB* genes. The sheep *DRB4* sequence was an outlier. Two pairs of genes were highly associated. The sheep *DRB5* gene and the *DRB* gene on clone c5-2 (Scott et al. 1991b) were grouped together in 100% of trials. They show only one amino acid difference from one another, both have a stop codon at the same position in exon 3,

Figure 4.10

Multiple alignment of the predicted amino acid translations of exons 2 and 3 from the *DRB* genes in cosmids 9.5 (*DRB1*), 9.1 (*DRB2* and *DRB3*), 23.1 (*DRB4*) and 46 (*DRB5*). + signs indicate positions in the antigen recognition site of the class ll molecule. * indicates a stop codon. The putative glycosylation site is in bold type.

exon 2

```
          10        20        30        40        50        60        70        80        90
           + + +             + +      ++                  +   +     +  ++  +   +   +   +
DRB1   HFLEYTKKECRFS**NGT**ERVRFLDRYFYNGEEYVRFDSDWGEYRAVAELGRPDAKYWNSQKEILERRRTEVDTYCRHNYGVIESFSVQRR
DRB2   L-MHQF-G-------L--M--FA--I--TK-D-H----V--FT-LT-------E---Q--DFM-QM-AK---V--S--QG-G--LR---
DRB4   L-MVQG-S--H-----QQ-------I--R--Q-----LV-----RT-M---A-ER--RWPDA-Q-A-AA-HDF-AS--EFFA-RT----
```

exon 3

```
          100       110       120       130       140       150       160       170       180
DRB1   EPIVTVYPAKTQPLQHHNLLVCSVNGFYPGHIEVRWFRNGHEEEAGVISTGLIQNGDWTFQTMVMLETVPQSGEVYTCQVDHPSRTSPITVEW
DRB2   --T--------------S-----------------------------------------V-------H---------E-------L----
DRB3   G-T--------------S-------------------------------------S---V-----L-H---------E-------L----
DRB4   --T-----V-SR--R-------------------K--Q--Q------V-----P-------I-----I---G------HVE-------V-A--
DRB5   --T-----E--------D------D---L--A-----WK-Q----------------*-L-------M------------E---Q---L---
```

and both have a non-consensus splice junction at the end of exon 3. Sheep genes *DRB2* and *DRB3* occurred together in 97.7% of trials. While the sheep *DRB2* and *Bota-DRB2* genes were not significantly associated in this analysis, they shared 92% identity in their amino acid sequence.

4.4.4 Repeat sequences in intron 2 of the sheep *DRB* genes

The region of intron 2 adjacent to the 3' end of exon 2 in *DRB* genes in many species including human, sheep, red deer and cattle, is characterised by the presence of simple tandem repeats (str), of basic structure $(gt)_n$ $(ga)_m$ (Reis et al. 1990, Sigurdardottir et al. 1992a, Ammer et al. 1992, Schwaiger et al. 1993a,b, Swarbrick et al. 1995). This is also true of the expressed *DRB1* gene in cosmid 9.5 (Ballingall et al. 1992). Inspection of the sequence of the *DRB4* pseudogene in cosmid 23.1 in this region shows that it also contains a repeat region but it is rudimentary compared with the usual str (Figure 4.7). There does however appear to be a minisatellite sequence derived from the basic element $ca(ga)_2$ in this region. The full length of this repeat region has not been sequenced and it has not been determined whether or not its length varies between individuals.

The equivalent region in the *DRB2* gene in cosmid 9.1 shows a slightly better preservation of the basic structure of the str, with an imperfect string of gt and ga motifs (Figure 4.7).

4.4.5 Polymorphism of the divergent sheep *DRB2* and *DRB4* genes

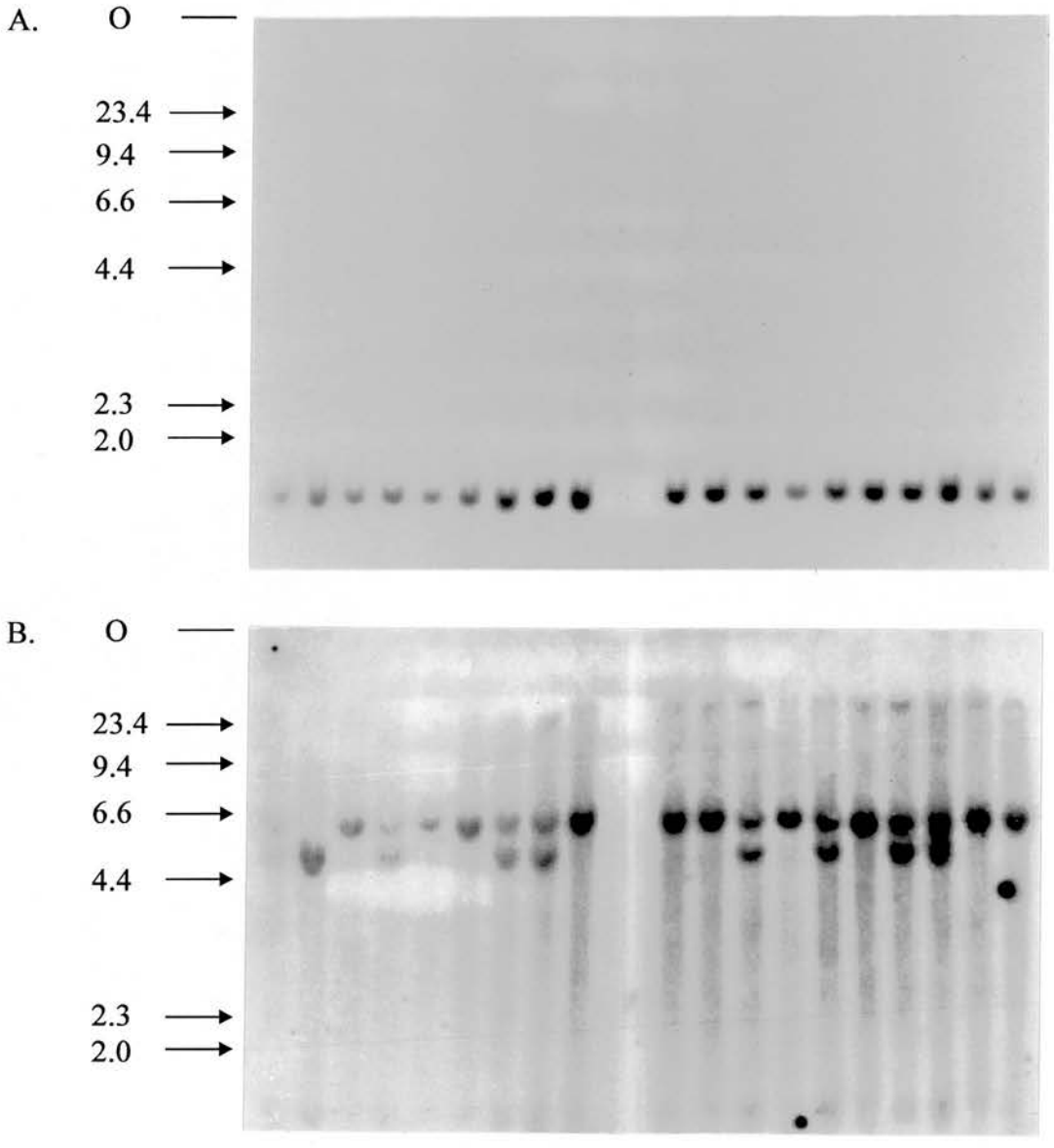
As it was initially thought that both the sheep pseudogene sequences *DRB2* and *DRB4* were novel (the *Bota-DRB2* gene sequence was not in the database), polymorphism at these loci was looked for in a group of twenty animals by RFLP using exon 2-specific probes from each gene. A 320 bp SacI/PstI probe containing most of exon 2 was obtained from the subclone of cosmid 23.1, and a 250 bp Sau3A1 fragment was obtained from the subclone of cosmid 9.1 (these restriction sites are

Figure 4.11

Southern blots of genomic sheep DNA digested with EcoRl and hybridised to

A, exon 2 probe from sheep *DRB2* and

B, exon 2 probe from *DRB4*. The blots were washed in 0.2xSSC, 0.1% SDS at 65^0C and autoradiographed for 3 nights at -70^0C.



indicated in Figure 4.7). 15 μg of genomic DNA from each sheep were digested with EcoRl and following agarose gel elctrophoresis, were blotted to nylon membranes and hybridised to the probes. The membranes were washed at high stringency and the resulting autoradiographs are shown in Figure 4.11. Clearly the *DRB2* gene was monomorphic for this enzyme, while the extra bands in some sheep provided evidence for the presence of at least two alleles at *DRB4*.

4.5 Discussion

In this chapter, the sequence of the sheep MHC class ll *DRA* gene is described together with the partial sequences of five *DRB* genes, one of which is the expressed *DRB1* gene (Ballingall et al. 1992).

Among the MHC genes, *HLA-DRA* and its mouse orthologue, *Eα,* are unique in being virtually monomorphic. Of four published *HLA-DRA* sequences, none show amino acid differences in the membrane distal domain (Korman et al. 1982, Larhammar et al. 1982, Lee et al. 1982b, and Das et al. 1983). Of six mouse *Eα* alleles, one differs at two amino acid residues from the rest (Ayane et al. 1986). These observations have now been extended to ruminants. The sheep sequence reported here differs at only two amino acid sites from the sheep *DRA* cDNA sequence reported by Fabb et al. (1993) and at only two sites from the *Bota-DRA* gene (van der Poel et al. 1990, Aida et al. 1994).

Phylogenetic trees place all of the *DRA* genes on a separate branch from *DQA* which suggests that the gene duplication which gave rise to these gene lineages predated mammalian speciation events (Klein and Figeroa 1986).

The minimal intra- and inter-species polymorphism in *DRA* suggests that some function of *DRA* has required extensive sequence conservation over a long period of mammalian evolution. In transgenic mice for example, *HLA-DRA* can replace *Ea* and behaves normally in antigen presentation and T cell repertoire selection (Lawrance et

al. 1989). Three possible explanations for this minimal polymorphism have been proposed. Firstly, it may be due to the requirement of *DRA* to pair with multiple *DRB* genes (Germain et al. 1985). Secondly, mutations in the *DRA* chain may be selected against because the *DR* isotype presents a unique set of peptides. Thirdly, *DR* may play a unique role in T cell repertoire selection. A strong counter-argument against all of these is the fact that 50% of MHC haplotypes in wild mice populations do not encode expressed *E* proteins (Figueroa et al. 1990) and so at least in mice, *DR* expression is not required for survival.

A problem which has confounded the mapping and characterisation of the *DQ* and *DR* genes of sheep class ll region in this thesis, is the fact that the cosmid libraries were constructed from DNA from three unrelated sheep which were heterozygous at the MHC. Given the haplotype variation in the number of *B* genes in the human *DR* sub-region (WHO Nomenclature Committee 1991), care needs to be taken in extrapolating from potentially six different haplotypes from three animals. However, in spite of this limitation, it is possible to draw some conclusions about the *DRB* genes in sheep 1 and 2.

There are clearly up to five different *DRB* genes in these sheep, one expressed locus, *DRB1*, two potentially expressible loci, *DRB2* and *DRB4*, one pseudogene, *DRB5* and a gene, *DRB3*, which was possibly truncated by the cloning process. The *DRB1* gene was cloned from sheep 1 and sheep 2. *DRB2* and *DRB3* were obtained from a cosmid cluster from sheep 1. Some of the restriction sites in this cluster were polymorphic (Figure 4.6) and so the clones probably represent two alleles at these loci. The *DRB2* gene was also obtained from sheep 2. *DRB4* was cloned only from sheep 1, while *DRB5* was cloned only from sheep 2. The occurrence of *DRB5* adjacent to the sheep *DRA* gene on one clone but not on the other is an analogous situation to that of the *HLA-DRB9* pseudogene which is adjacent to *HLA-DRA* in some but not all haplotypes in the human (WHO Nomenclature Committee 1991). However, *HLA-DRB9* is an isolated β1 domain exon (Meunier et al. 1986) rather than β2 as in the sheep.

The *DR* sub-region of sheep 1 therefore contains at least four *B* genes, *DRB1-4.* and depending upon the haplotype, it may also contain *DRB5*. The *DR* sub-region of sheep 2 contains at least 3, *DRB1*, *2*, and 5. Southern blots would indicate that DRB2 and 4 are ubiquitous The sheep *DR* sub-region has an extended *HLA-DR*-like structure rather than the smaller *I-E* sub-region of the mouse. Unfortunately, we have not been able to link the various clones together. However, if we allow that sheep 1 has a *DRA* gene, then we can estimate the minimum length of the sheep *DR* sub-region to be approximately 170 kbp.

The sheep genes *DRB2* and *DRB5* show very high homology to the *Bota-DRB2* gene (Muggli-Cockett and Stone 1989) and a sheep *DRB* pseudogene (Scott et al. 1991b). The sheep *DRB4* gene on the other hand is quite divergent. The sequence of the third exon of this gene shows that it is clearly *DRB*-like.

The second exons of a large number of alleles at the expressed *DRB1* locus in sheep, cattle (*DRB3),* goats and red deer have now been sequenced (Sigurdardottir et al. 1991, Ammer et al. 1992, Schwaiger et al. 1993a,b, Swarbrick et al. 1995). Downstream of the second exons of expressed *DRB* genes in man (Andersson et al. 1987, Reis et al. 1990), mouse (Braunstein and Germain 1986), cattle (Muggli-Cockett and Stone 1988, 1989, Groenen et al. 1990), sheep (Schwaiger et al. 1993) and red deer (Swarbrick et al. 1995), there occurs a simple tandem repeat (str) of general formula $(gt)_n(ga)_m$. The length of this str has been shown to be polymorphic (Reis et al. 1990) and has been used as a method of typing sheep and cattle by estimating the length of the str in different individuals following PCR amplification (Schwaiger et al. 1993, Ellegren et al. 1993). Variation in the str has been shown to correlate with the sequence of exon 2 which encodes the β-pleated sheets, as opposed to that which encodes the α-helices of the peptide binding groove (Ammer et al. 1992, Sigurdardottir et al. 1992a, Schwaiger et al. 1993a,1994). It would appear that these two regions of the β1 domain have different evolutionary histories.

The strs of the pseudogenes *DRB2* and *DRB4* described here are now only rudimentary fragments. That of *DRB2* is somewhat better preserved and this may indicate that this gene was expressed for sometime after its generation by gene duplication (Schwaiger, personal communication).

The sheep *DRA* and *DRB* genes described here express a *DR* molecule on the surface of the mouse L cell following transfection, and the transfected cell line has been used to help define the specificities of anti-sheep class ll mabs (Ballingall et al. 1992, 1995). The transfected L cell line has also been used to demonstrate that two distinct *DR* molecules are expressed in sheep based on the reactivity of *DRB*-specific mabs (Dutia et al. 1994). Unfortunately, the amino terminal of the β chain which was immunoprecipitated by the mabs was blocked and so no amino acid sequence was obtained to compare with the sequences described here. The status of the *DRB* genes *DRB2*, *3* and *4* remains uncertain. In cattle, the level of transcription of the *Bota-DRB3* gene was estimated to be 200 times that of *Bota-DRB2* (Burke et al. 1991).

CHAPTER 5

The Sheep MHC class II *DNA* and *DOB* genes

5.1 Introduction

The nucleotide sequence and genomic structure of a human MHC class ll *A* gene distinct from *DP, DQ* and *DR* genes was described by Trowsdale and Kelly (1985). The gene was designated *DZA*, but has since been renamed *DNA* (WHO Nomenclature Committee for factors of the HLA system, 1991). The *HLA-DNA* gene is as distantly related to the genes of the *HLA-DP, DQ* and *DR* sub-regions as these genes are to themselves. When a molecular map of the HLA-D region became available, it was shown that the *HLA-DNA* gene was unusual in not having a *B* gene partner situated within a few kilobases, the nearest *B* gene being the non-polymorphic *HLA-DOB* gene (Tonnelle et al. 1985, Servenius et al. 1987) which is situated 160 kbp telomeric to the *HLA-DNA* gene (Campbell and Trowsdale 1993).

In the mouse, the equivalent genes, *H-2Oa* and *H-2Ob* (formerly Aβ2, Larhammar et al. 1985), map 150 kbp apart (Karlsson and Peterson 1992). However, the mouse genes express an MHC class ll protein of unknown function whose tissue distribution is restricted to B cells and epithelial cells of the thymic medulla (Karlsson et al. 1991, Karlsson and Peterson 1992). Although no *DN*-like protein product has been described from other species, *DNA* and *DOB* genes are clearly expressed in some context. cDNAs for the *DNA* gene have been described from the human (Trowsdale and Kelly 1985, Johnsson and Rask 1989, Young and Trowsdale 1990), mouse (Karlsson and Peterson 1992) and tammar wallaby (Slade et al. 1993), while *DOB* cDNAs have been described from the human (Tonnelle et al. 1985), mouse (Larhammar et al. 1985), rabbit (Chouchane et al. 1993) and chimpanzee (Kasahara et al. 1989). Interestingly, Stone and Muggli-Cockett (1993) were unable to detect *DOB* transcripts in RNA from cattle peripheral blood lymphocytes using the heterologous *HLA-DOB* cDNA probe.

The presence of *DNA* and *DOB*-like genes in cattle and sheep was inferred from Southern blots of genomic DNA using *HLA* gene probes (Scott et al. 1987, Andersson et al. 1988, Trowsdale et al. 1989). However, the only *DOB*-like

sequences in the nucleotide databases are those of the human, mouse, rabbit and chimpanzee as described above, and the only *DNA*-like sequences in the database are those of the human, mouse and tammar wallaby.

The strong sequence conservation between orthologous genes from evolutionary disparate groups argues that these genes produce functional products. The detection, cloning, sequencing and expression of the sheep MHC class ll *DNA* and *DOB* genes are described in this chapter, extending the study of the *DN/DO* region to a member of the ungulate family,

5.2 The *Ovar-DNA* gene.

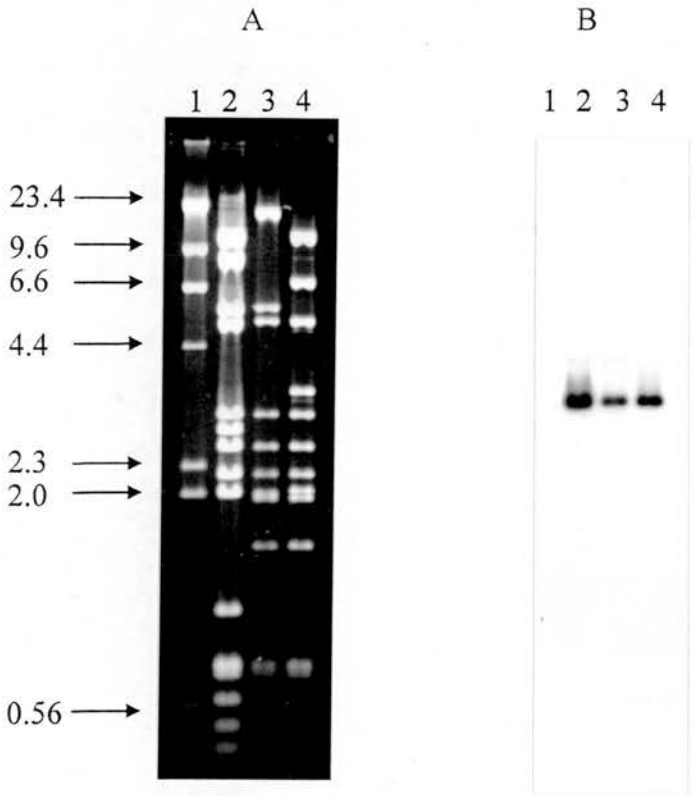
Inspection of the restriction digests of cosmid 12 from sheep 1, and cosmids 24.1 and 46.1 from sheep 2, indicated that these three clones were closely related (Figure 5.1a). The 3.2 kb EcoR1 fragment shared by all three clones hybridised to a selection of human *A* gene probes (Figure 5.1b).

Restriction maps of the cosmids using the enzymes EcoR1, BamH1, Hindlll, Sac1 and Sma1 were constructed and the individual cosmids aligned as shown in Figure 5.2. There were a limited number of differences in restriction sites between the clones from the two sheep. Four sites in cosmids 24.1 and 46.1 were not represented in cosmid 12, while five sites in cosmid 12 were not shared by cosmids 24.1 and 46.1, as shown in Figure 5.2. The position of the 3.2 kb EcoR1 hybridising fragment is indicated by a solid box in Figure 5.2. This fragment was excised from an agarose gel, the DNA purified and ligated into the pBS+ vector. Recombinants were sequenced from the forward and reverse primers of the pBS+ vector. The sequence data obtained were compared with those in the EMBL database using the FASTA comparison program, but showed no homology to class ll genes. A restriction map of the subclone was produced by digesting the DNA with enzymes from within the multiple cloning site of pBS+. The map is shown in Figure 5.3. Further subclones were produced in the bacteriophage vector M13mp18/19 to obtain ssDNA to

Figure 5.1

A. EcoRl restriction digests of the cosmids containing the Ovar-DNA gene.

B. Blot of the gel in a. hybridised to HLA-DQA probe, showing the 3.2 kbp EcoRl fragment which was sequenced. The blot was washed at 42^0C in 0.2xSSC, 0.1% SDS and exposed overnight at -70^0C for autoradiography.



Track 1 λ HindIII standards

Track 2 cosmid 12

Track 3 cosmid 24.1

Track 4 cosmid 46.1

facilitate subsequent sequencing. These subclones contained the fragments produced by EcoR1/Kpn1, Kpn1/Sma1, Sma1/Hindlll, and Hindlll/EcoR1 double digests, reading from the left hand side of the EcoR1 clone as drawn in Figure 5.3. Sequence data was obtained from these subclones from the M13 primer as indicated by the arrows in Figure 5.3. A number of sequencing primers (17mers) were constructed to confirm the sequence in some areas and to sequence the complementary strand. As sequence information accumulated, it became clear from database searches that the gene was the sheep orthologue of the *HLA-DNA* gene. This was the first report of the sequence of an ungulate MHC class ll *DNA* gene (Wright et al. 1995).

The sequence of the sheep *DNA* gene is given in Figure 5.4. The *Ovar-DNA* gene has all the salient features of an MHC class ll *A* gene. Exons two and three code for the two extracellular domains of the protein. Exon 4 codes for a proline-rich connecting peptide, a hydrophobic transmembrane region and all but the third nucleotide of the last codon of the cytoplasmic tail. A fifth exon contains the last nucleotide of the cytoplasmic tail, two in-frame stop codons and the 3' untranslated region. A putative, non-consensus polyadenylation site, AGTAAA, is indicated in bold type at position 1823. The gene also carries two conserved N-linked glycosylation sites NGT and NAT, indicated in italics in Figure 5.4, and the two conserved cysteine residues which form the disulphide bond in the alpha-2 domain..

The predicted amino acid translations of exons 2-4 of the published *DNA* genes are compared with the sheep sequence in Figure 5.5. Table 5.1 shows the percentage homology at amino acid level between exons 2-4 of the sheep gene and the other published *DNA* genes. The sheep sequence was most like that of the *HLA-DNA* gene with an amino acid identity at the second exon of 78%. This identity fell to 76, 53, 54, 42 and 53% when the second exon of the sheep gene was compared to those of the mouse gene, *H-2Oa,* and the other sheep class ll *A* genes, *Ovar-DQA1, DQA2, DRA* and *DYA*, respectively. The sheep and human genes shared 83% nucleotide

Figure 5.2

Restriction maps of the cosmids 12, 24.1 and 46.1 which contain the *Ovar-DNA* gene. The closed box indicates the position of the gene, while the open box indicates the position of the SINE sequence as described in the text. Open circles indicate sites which are unique to cosmids 24.1 and 46.1, while the closed circles indicate sites unique to cosmid 12. The arrow indictes the direction of transcription of the gene.

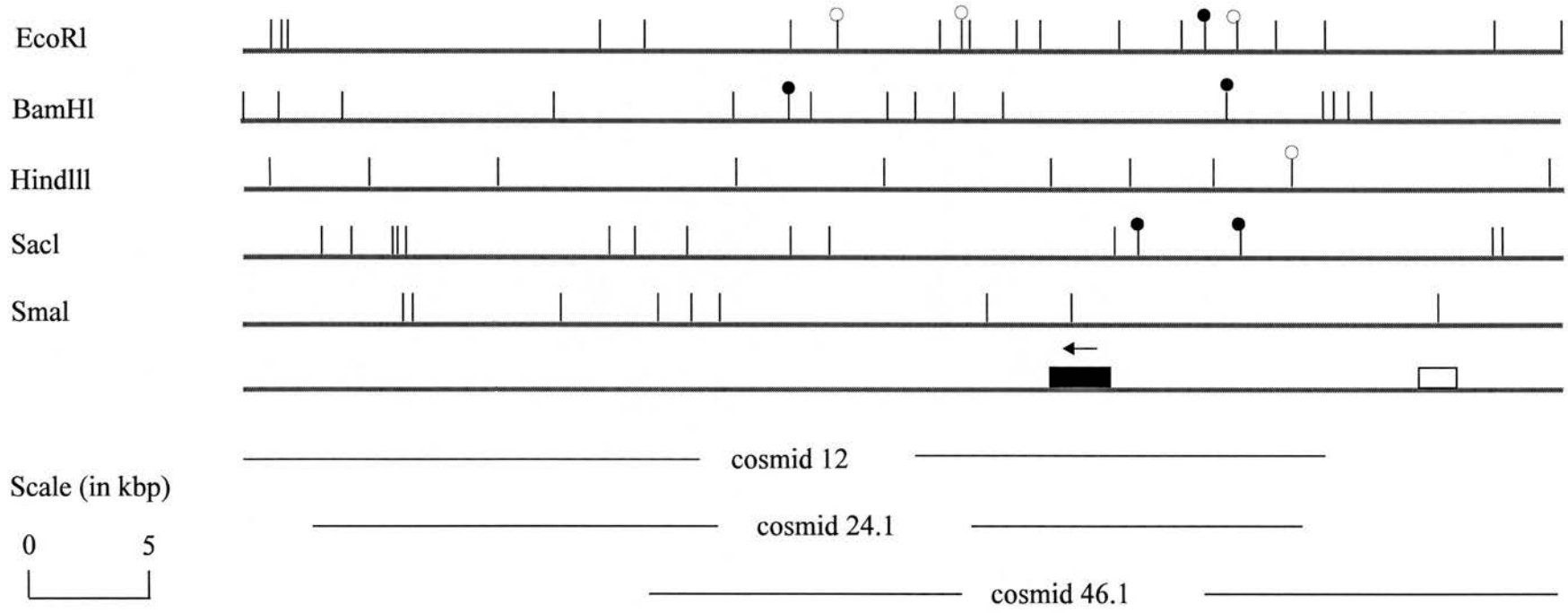


Figure 5.3

Restriction map of the 3.2 kbp sub-clone of cosmid 12 which contains the *Ovar-DNA* gene. Arrows indicate the region sequenced. The black boxes indicate the positions of the coding regions. The lightly shaded area of the cosmid was not sequenced.

Abbreviations: CP connecting peptide, TM transmembrane domain, CT cytoplasmic tail, 3' UT 3' untranslated region.

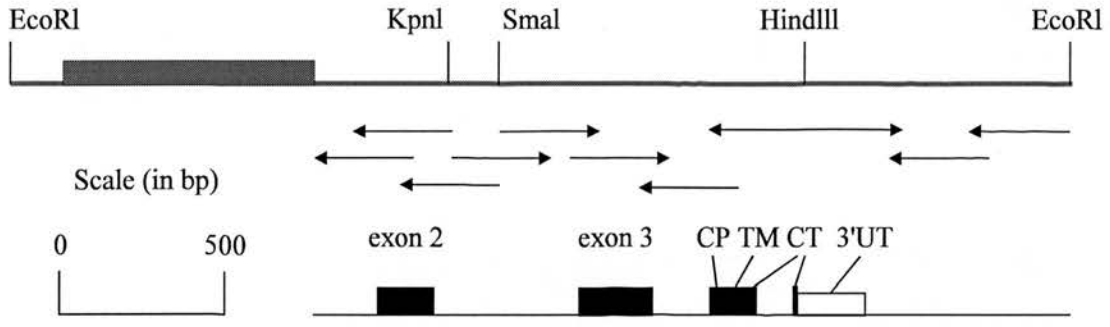


Figure 5.4

The nucleotide and predicted amino acid sequences of the *Ovar-DNA* gene. Coding sequences are in *capital letters* with the translation above each codon in the single letter amino acid code. Splice junctions are *underlined* and the putative polyadenylation signal is in *bold type*. Positions of the primers DN1-3 are indicated.

```
ttccaaccctcagcaactcctctgtctccagggaccaactcagtgtcagaacaggactct   60
tcccccacacagagcccaggatggggcaatagctcagaactcaacccatcacacctgatg   120
                                                        exon 2
                                                          D
gatgagagctgataccttttcttttctcccccctgaccctgaactcacccaacacagCTGA   180

  H  M  G  S  Y  G  P  A  F  Y  Q  S  Y  D  G  A  G  Q  F  T
CCACATGGGCTCCTATGGACCAGCCTTCTATCAGTCCTATGACGGCGCAGGTCAGTTCAC   240
          5' atggaccagccttctat 3' (primer DN1)

  Y  D  F  D  G  E  Q  L  F  S  V  D  L  K  K  R  E  A  V  W
CTATGATTTTGATGGAGAACAGCTGTTCTCTGTGGACCTGAAGAAGAGGGAGGCTGTGTG   300

  R  L  P  E  F  G  N  F  A  Y  F  D  P  Q  N  G  L  V  S  I
GCGTCTGCCTGAATTTGGCAACTTCGCCTACTTCGACCCGCAGAACGGGCTGGTCAGTAT   360

  A  M  I  K  A  H  L  E  V  L  V  E  R  E  N  G  T  R  A  P
CGCGATGATCAAAGCGCATCTGGAGGTCTTGGTGGAGCGCTCCAATGGCACCAGAGCCCC   420

  N
CAACGgtaccgggcccttcccttgcccacccccccccccaggcccccagcaagcagggg   480

aggtccagagaagcctctgccccacccctcaggcctctgttggctccctggccccactct   540

caaaatctgagactgttgttctcaacccctttctttcccgggaggttctagatctcctca   600

gattccttccacagccccttcctataggaatgacccctcctgcctggcagatacggcacc   660

ccctgggcacccccctgtcccttcattctcttgtccttctgttccccgcttctgccccagt   720

gtcaccagcccagagcctccactgccctcagaaggctaataccaggaagctccctcaatc   780

catcctgatgaggccccagctggaggctggtcctgtgggattcaaggcctgcttcttcct   840

  exon 3
  V  P  P  R  V  A  V  L  P  K  S  R  V  Q  L  G  E  P  N  V
agTGCCCCCCAGAGTGGCTGTGCTGCCCAAGTCTCGCGTGCAGCTGGGCGAGCCCAACGT   900

  L  I  C  I  V  D  N  I  F  P  P  V  I  N  A  T  W  L  H  N
CCTCATCTGCATCGTGGACAACATCTTCCCCCCGGTCATCAACGCCACCTGGCTGCACAA   960
                                                      3' acaa
```

```
  G  Q  P  V  T  Q  G  V  T  Q  T  S  F  Y  A  Q  P  D  H  S
CGGCCAGCCCGTCACCCAGGGCGTGACCCAGACCAGCTTCTACGCCCAGCCTGACCACTC   1020
cggccagcccgtc 5'  (primer DN2)

  F  R  K  F  H  Y  L  T  F  V  P  L  A  D  D  F  Y  D  C  K
GTTCCGGAAGTTCCACTACTTGACCTTTGTGCCCTTGGCGGACGACTTCTATGACTGCAA   1080

  V  E  H  W  G  L  D  Q  P  L  F  Q  H  W
GGTGGAGCACTGGGGCCTGGACCAGCCGCTCTTTCAGCACTGGGgtacggagcccctcc   1140

ctgctgcccacctcagacccatgcatccttcaccctgaatctctttgcaggacccctga   1200
                    exon 4
                    E  P  Q  V  P  T  V  L  P  D  T  T  E  T
ccttccctttctctcccagAGCCCCAGGTGCCCACCGTGCTGCCTGACACCACAGAGACT   1260

 L  V  C  A  L  G  L  I  L  G  L  G  G  F  L  G  G  I  T  L
CTGGTCTGTGCCCTTGGCCTGATCCTCGGCCTGGGGGGCTTCCTCGGGGGCATCACCCTC   1320

 I  I  T  G  T  C  L  S  S  A  P  R
ATCATCACAGGCACGTGCCTGTCCAGCGCCCCCAGgtgcattgggggtctcgggggagga   1380
       5'gcacgtgcctgtccagc 3'  (primer DN3)

aggtgggactctgcacaggatgggggtggaggacgagagattctgggaggggagcaggga   1440

caagcagaggggaggcaaggggtcaggttggagggaggtgggcagagaaggacccagaa   1500
                                                    stop
ttgtgagcttgaagcttgatcacgtggtttttgctactttagGtgatgacccttctgaga   1560

gaagcaagtgtgagacaccattcagctccctggcaagtttctgacggcccttgcctgctc   1620

aatgtgactccaccaagttcatctatgccgactttgaatgggatcaacccctgtcccata   1680

ggtctccttttttggcccccatactcatggcagggacttgtggggaaccaacacgctccct   1740

ttcccagccccaacacacacacacacagacacacacacacacacacacactttattggtt   1800

tacccaaagctctggccagcacagtaaaaattttaacaatgtttgcacagtgtccatcttc   1860

ccaggccttagccaattctctaagggtgtaagggacactaggggcctcccaggtggctca   1920

gtgggtaaagaatgtgcctgtcatgcaggagatatgggtttgatccctgaatcgggaaga   1980

tcccctggagaagggcatggcaaccacttcagtattcatgcctggagaaccccatgggca   2040

gaggagcctggtgggctacagtccgtggggtcccaaagagtcggatatgactgaagggac   2100

tgagcctccgtgcaaaggactccagggtctggaagaggggctttctggtgatggccactg   2160

agcgtgacgtgctgctggggactggcgttcattcacgggccaatctagctatgtcttgct   2220

tcatttgacttgtcagcgtggggccccttgtgtgtcacttgtgaacggaccacttcacat   2280

ggacgtcataaatttatatcctgaattc                                   2308
```

Figure 5.5

Multiple alignment of the amino acid translations of exons 2-4 of the human, sheep, mouse and tammar wallaby (*Macropus eugenii*) *DNA* genes.

CP is connecting peptide, TM transmembrane domain, CT cytoplasmic tail. Maeu *Macropus eugenii* (tammar wallaby).

```
          exon 2
HLA-DNA   ADHMGSYGPAFYQSYGASGQFTHEFDEEQLFSVDLKKSEAVWRLPEFGDFARFDPQGGLAGIAAIKAHLDILVERSNRSRAIN
Ovar-DNA   --------------DGA----YD--G----------R----------N--Y----N--VS--M-----EV------GT--P-
H-2OA     ---------------D----------G--I------NE-V----------HS-F-S--MS-SM--------------T--VS
Maeu-DNA                              -.--Y---Q-K-T-------SHSSS-------REK-TC-YN----IK--------S

          exon 3
HLA-DNA   VPPRVTVLPKSRVELGQPNILICIVDNIFPPVINITWLRNGQTVTEGVAQTSFYSQPDHLFRKFHYLPFVPSAEDVYDCQVEHWGLDAPLLRHW
Ovar-DNA  -----A-------Q--E--V--------------A---H---P--Q--T-----A----S-------T---L-D-F---K-------Q--FQ--
H-2OA     ----T-----K--V------D-------V-----S-PI-K----------N-R-------T----------K-------T---Q--
Maeu-DNA  --E---FSE-P-------V---L--------V--K------VI-T--SE-D---R---K----Y--T-L-NT--F-

          CP             TM                          CT
HLA-DNA   ELQVPIPPPDAM   ETLVCALGLAIGLVGFLVGTVLII    MGTYVSSVPR
Ovar-DNA  -P---TVL--TT   ---------IL--G---G-IT---    T--CL--A-
H-2OA     -P--LT----TT   ---I-G---VL--M-C-L----M-    T--RRP-IR-
```

Table 5.1. The % identity between the predicted amino acid sequences of exon 2, exon 3, connecting peptide, transmembrane domain and the cytoplasmic domain of the *Ovar-DNA* gene and those of the human (Trowsdale and Kelly, 1985,), mouse (Karlsson and Peterson, 1992) and tammarin wallaby, (*Macropus eugenii*), (Slade et al. unpublished 1993, Acc. No. l12121) *DNA* genes, together with those of an *Ovar-DQA1* (Fabb et al. 1993, Acc. No. m93430), *DQA2* (Fabb et al. 1993, Acc. No. m93431), *DRA* (Ballingall et al. 1992, Acc. No. z11520) and *DYA* (Wright et al. 1994, Acc. Nos. z27398 and z27399) gene. * Only partial sequence available. Figures in parenthesis are the % similarities. Abbreviations : CP, connecting peptide, TM, transmembrane domain and CT, cytoplasmic tail.

	HLA-DNA	*H-2Oa*	Maeu-DNA	Ovar-DQA1	Ovar-DQA2	Ovar-DRA	Ovar-DYA
Exon 2	78 (88)	76 (84)	54 (70)*	53 (70)	54 (71)	42 (71)	53 (72)
Exon 3	80 (88)	80 (87)	66 (80)*	64 (76)	66 (78)	64 (78)	71 (79)
CP	54 (54)	77 (77)	*	38 (69)	38 (69)	15 (54)	*
TM	74 (78)	61 (78)	*	56 (74)	57 (78)	*	*
CT	60 (80)	50 (50)	*	30 (40)	33 (44)	*	*

identity at exons 2 and 3. This figure agreed well with the 82% identity for the human versus cattle comparison (Andersson and Gustafsson, quoted as a personal communication in Rosen-Bronson and Long, 1991).

5.3 The *Ovar-DOB* gene

Cosmid clone number 305 which hybridised weakly to *Ovar-DQB* and *Ovar-DRB* gene probes was obtained from the pCos8 cosmid library from sheep 3. EcoRl, BamHl, Hindlll, Sac1 and Sma1 digests of the clone are shown in Figure 5.6. The clone was mapped using these five enzymes and shown to contain a single *B* gene whose position is indicated by the closed box in Figure 5.7. There was a relative paucity of convenient restriction sites in the region of the gene and consequently the first sub-clone that was obtained was the relatively large 5.5 kbp Hindlll/EcoRl fragment. A restriction map of this clone is shown in Figure 5.8. The subclone was digested with a number of enzymes, blotted and hybridised to an *Ovar-DQB* exon 3 probe at low stringency, two Pstl fragments were positive indicating that there was a Pstl site in the third exon of this gene. The cosmid was cut with Pstl and the 1-2 kbp fragments shotgun cloned into pBluescript(SK+/-). Fortuitously, three Pstl clones were obtained, one of which contained exon 1 (mp1), another contained exon 2 plus the 5' end of exon 3 (mp2) and the third contained the rest of exon 3 and downstream sequences (mp3). The insert in mp1 contained the Hindlll site from the original Hindlll/EcoRl fragment (Figure 5.8). The Pstl/Hindlll fragment was excised and cloned into pBluescript(SK+) and (KS-) and sequenced. Exolll deletion libraries were prepared from the other two Pstl clones and some useful sequence obtained. However, problems with compressions necessitated reverting to the ssDNA rescue system and walking along the sequence using puchased primers.

The nucleotide sequences of the regions which contained exons 1, 2 and 3 are given in Figure 5.9 together with the predicted amino acid translations. Within the region proximal to the promoter, the sheep and human sequences contained the same variation, TTCCAATCC (double underlined in Figure 5.9), on the consensus

Figure 5.6

Restriction fragments produced by digestion of cosmid 305 with EcoRl (E), BamHl (B), Hindlll (H), Sacl (Sa) and Smal (Sm). The positions of the 1 kbp molecular weight markers (M) are indicated by arrows. The units are kbp.

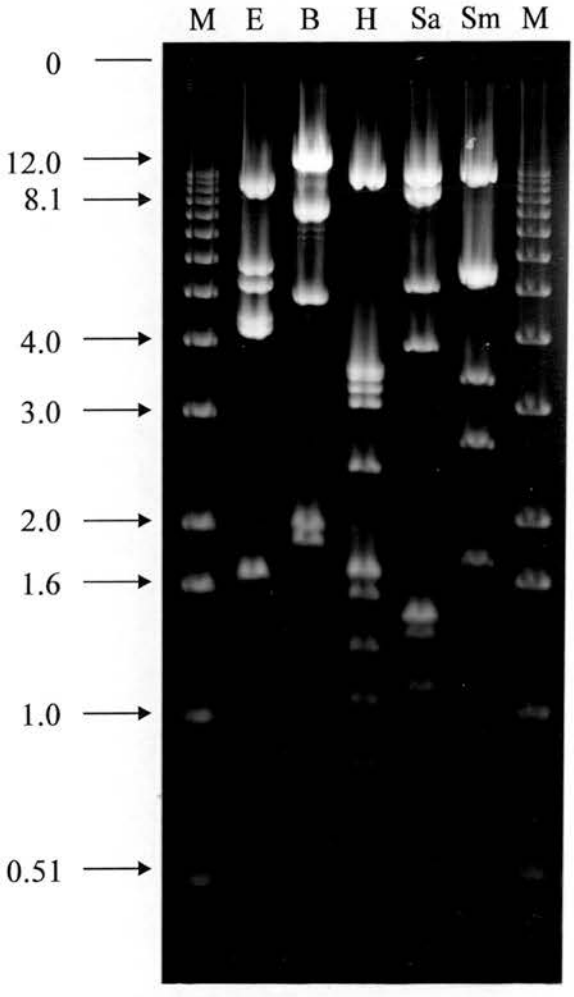


Figure 5.7

Restriction map of cosmid 305 for the enzymes EcoR1, BamH1, Hindlll, Sac1 and Sma1. The position of the *Ovar-DOB* gene is indicated by the black box and the arrow indicates the direction of transcription.

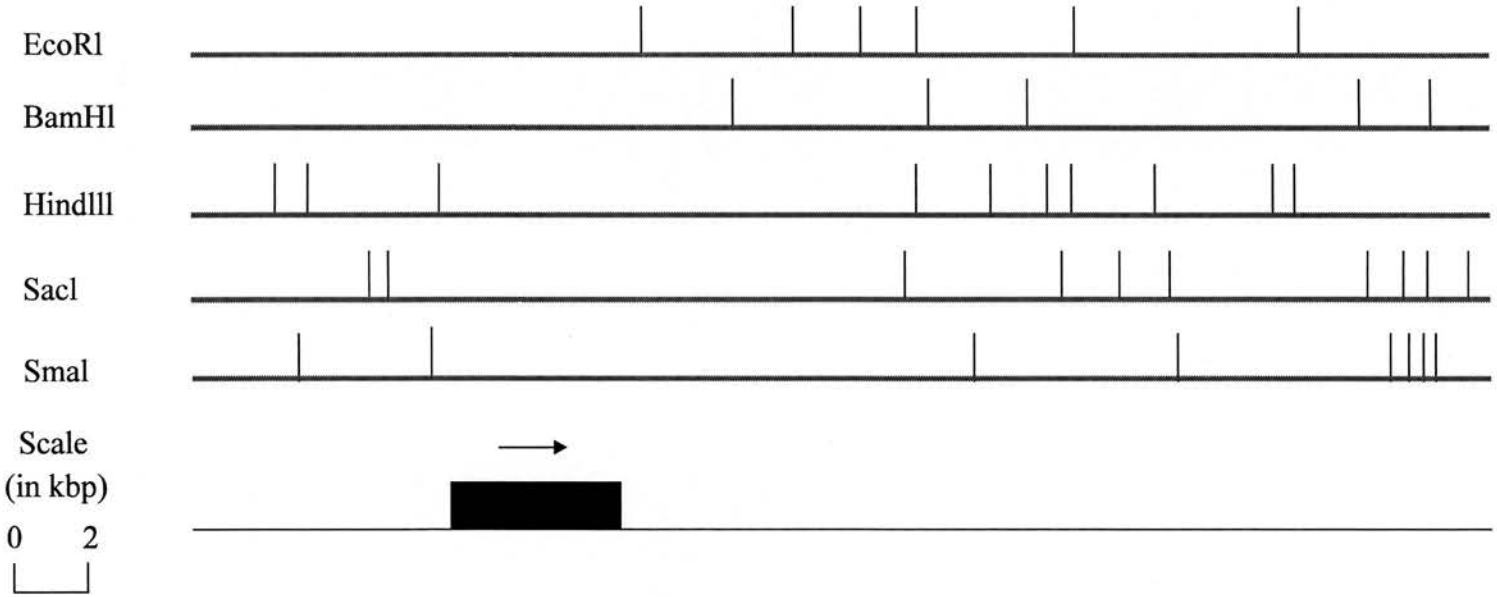


Figure 5.8

Restriction map of the subclones containing the *Ovar-DOB* gene. Arrows indicate sequencing strategy. Exons are marked by the closed boxes. The open box shows the area of high homology to the Bov2 SINE sequence. See text for details.

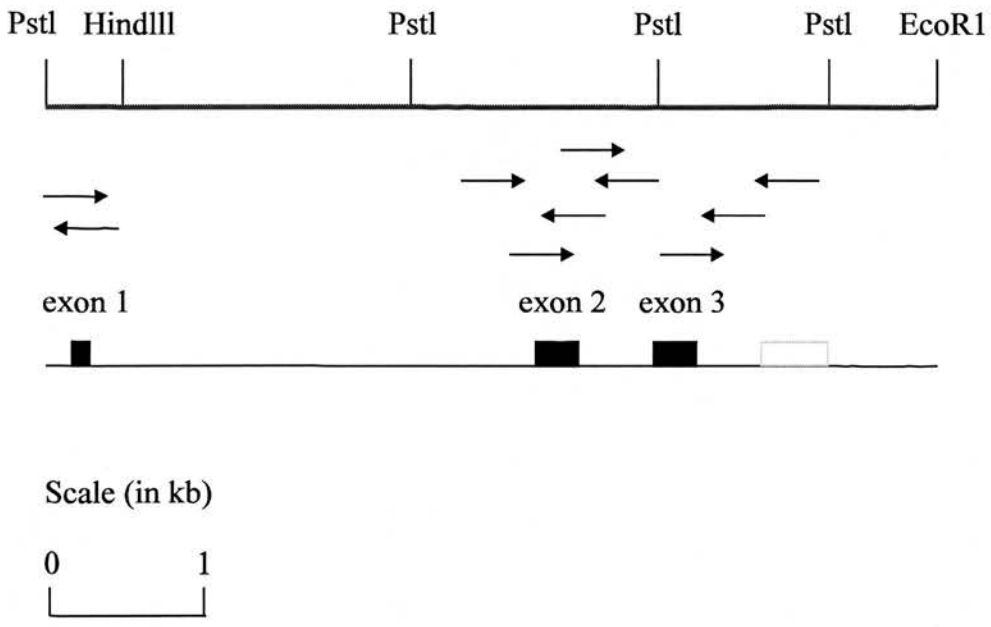


Figure 5.9

The nucleotide sequence and predicted amino acid translation of the *Ovar-DOB* gene. Coding sequences are in *capital letters* with the translation above each codon in the single letter amino acid code. Splice junctions are *underlined.* The sequence motifs involved in the regulation of transcription are *double underlined.* A putative N-linked glycosylation site is indicated in *bold.*

```
                                        S box                    X1 box
ctgcaggtgaaaaattctggagtggtatcacatgactttatgaaggcatcctacccagca  60

                                Y box
cctaatgacactaagctcctctagcttgattggttatctggatgaggagtttccaatccc 120

ggggtagacagaaagacatctctgctctgcatgaatatttgactgagtgtgtgtggccaa 180

                                                        exon 1
                                                        M  S  P
gcccattctaaagaggtggagatttcgtgtcttgtttgctttcttccagaATGAGTCCTA 240

I  W  V  P  W  V  V  A  F  S  S  T  V  L  R  L  D  A  S  M
TCTGGGTTCCCTGGGTGGTGGCCTTCTCATCCACGGTTCTCAGGCTGGATGCTTCTATGA 300

T  Q  G  R  D  S  P
CTCAAGGCAGAGATTCTCCAGgtaagaagatagtgatttctttcaggtatagaggcacac 360

attggcacgggggagagaggcctttccctacaggtccaggttccagtccagacaggaagc 420

agctctgacttcagagaatctttctcagagaaaagctt ..................... 458

.....1.5 kbp of intron 1 not sequenced.....

atggacagacaccacctaagcgactgaactgcatattccattgtatatgtgtaccacttt  60

ttttttttttaaagaatagatgtctgtttgctgagaggaaccagggtgtggcccagatg 120

agcagaggctactttgtcagatttactgaggtcaaagccctcctctccaatccatatggt 180

gcattgtaatataccaggtaggggtttcctaaagatgggggctggttcatagtcttttt 240

          exon 2
          E  D  F  V  T  Q  A  K  A  D  C  Y  F  T  N  G  T
ggttttccagAAGATTTTGTGACCCAGGCAAAGGCTGACTGTTACTTCACCAATGGGACA 300

 E  K  V  R  F  V  V  R  F  I  F  N  L  E  E  Y  A  R  F  D
GAAAAAGTGCGGTTTGTGGTCAGATTCATCTTCAACCTGGAGGAGTATGCACGTTTCGAC 360

 S  D  L  G  M  F  V  A  L  T  E  L  G  K  P  D  A  E  L  W
AGCGACTTGGGAATGTTTGTGGCCTTGACGGAGCTGGGGAAGCCCGATGCTGAGCTGTGG 420

 N  N  R  P  D  I  L  A  R  S  R  A  S  V  D  M  L  C  R  R
AACAATCGGCCGGATATTCTGGCGAGGAGTAGAGCCTCTGTGGACATGCTTTGCAGACGC 480
 N  Y  K  L  G  A  P  F  T  V  G  R  R
AACTACAAGCTGGGTGCACCCTTCACCGTGGGGAGGAGAGgtgaggtgggagttggggtc 540
```

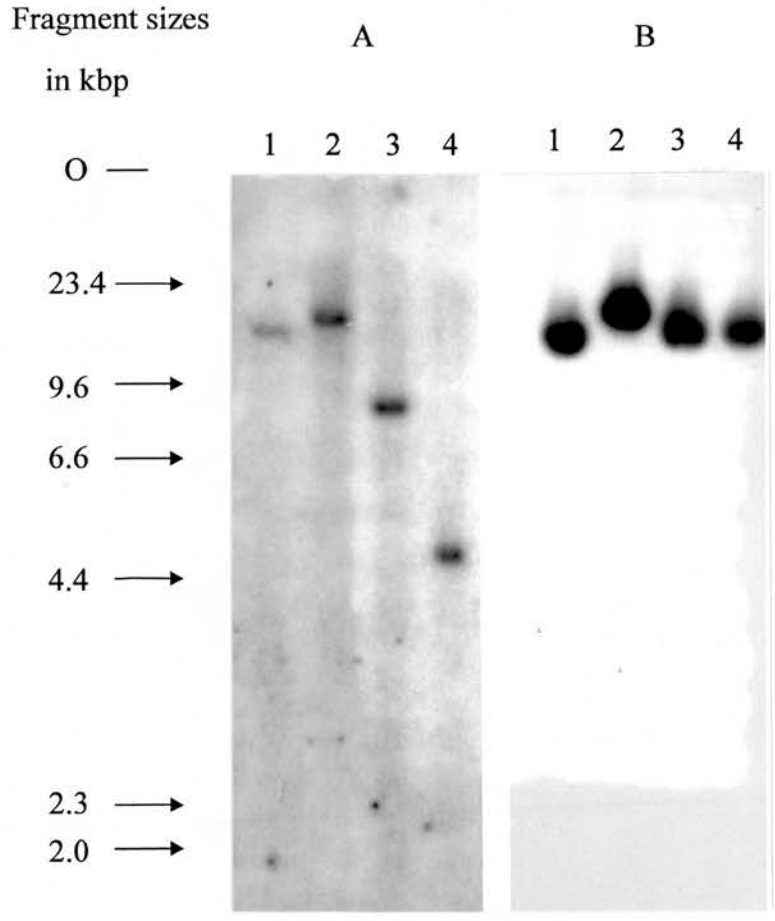
```
tcctaggtgggagcctgggtgtggctcttcttccttgtccgggcagtttcctgcttcagg 600
atatctatccatcagggtcaggtgcaggggagactatatatgagctgtccccagttgtga 660
ctcactctcactgtcctcaggggtcaggtcactgtgatcttggtcccagataccagttct 720
tggcggcctttgagggctagtgtcttcctgaacacaaggaaaattgtaggtgtgtctgat 780
aaagcccacatcctccttgtggacactgacatttccagtgtctgagtcctgggggtgagc 840
tttccccacatttccggttgccaggtgggtgatctggggaagagactcagggcagagaca 900
gcccctaaaggcactccggtgttggctttgtggatttccttcaccacctctcatttttc 960

          exon 3
            V  Q  P  E  V  T  V  Y  P  E  K  T  P  A  L  Q  H
tctcctcctagTGCAACCAGAGGTGACAGTGTATCCGGAGAAGACCCCAGCCCTGCAGCA 1020
  R  N  L  L  L  C  L  V  T  G  F  Y  P  G  D  I  K  V  T  W
CCGCAATCTGCTGCTTTGCTTGGTGACAGGTTTCTACCCAGGGGACATCAAGGTCACCTG 1080
  F  R  N  G  Q  E  Q  R  E  G  V  M  S  T  G  L  I  R  N  G
GTTCCGGAATGGGCAGGAACAAAGAGAGGGGGTCATGTCCACCGGCCTCATCAGGAATGG 1140
  D  W  T  F  Q  T  T  V  M  L  P  M  T  P  E  L  G  E  V  Y
AGACTGGACCTTTCAGACGACGGTGATGCTGCCCATGACGCCTGAGCTTGGAGAGGTCTA 1200
  T  C  L  V  D  H  P  S  L  P  S  P  V  S  V  E  W
CACCTGCCTTGTTGACCATCCCAGCTTGCCGAGCCCTGTTTCTGTGGAGTGGAgtgagaa 1260
ccagcttgtacttctgaacccaacttgggtcgatctgaccctggacagggcctgtgttac 1320
ctccgtattcaccttggccacagtactagttcctccctctgacctgaaggagtcatagaa 1380
attgggttcgttgggctgggttggagaaaatcatggcagatagctactcttggatcacat 1440
gtatgtcaaacatgcactgcttggctcgataaataaaaattactaatacaatgtacaaag 1500
aatccttcactgaaattaatgcttcttgagtttactccaagttatatcacatcttgtttt 1560
cagttctgttgaacacattgtgactaccatcccaccactggaaatcatgatcatgatttc 1620
cagctatggttcctttgaaggggctcagatctgagaactgcgagcaccgtggacagcacc 1680
tggggtcactctaactccctagagggaggcgagggcagatgctgacaaggacctcagggt 1740
gaatgagaaagaagtcatagcaggagacaggttctcatgaatgtgctgtaaatacaccat 1800
gacttttttttttgcaattagagagaatgcagtaaatgcagtctttaaatatagtcaaa 1860
gctatggtttttcaatagtcatgtacaggtgtgagagctgggccataaagaacgctgag 1920
cactaaagactgatgctttcaaattgcagtgcaggagaagactcttcagggtcccttgaa 1980
ctgcaagaaaattaaaccagtcaatccgaaaagaaaccaaccctgatattcattggaagg 2040
actgatgctgaagctgaagctccgatactttggtcacttgacatgaagagctgactcact 2100
ggaaaagaccctgatgctgggaaagaccaagggcaaaggagaaggggcagcagaggatg 2160
agatggttggatagcactgccaactcaatggacaccaatttgagcaagctctggcagaca 2220
gggaaggacaggggagcctggcatgctgcag                               2251
```

eucaryotic CAAT box, GG(C or T)CAATCT, 120 bp upstream of the initiation codon. The X, Y and S boxes involved in the regulation of MHC class ll gene transcription (Benoist and Mathis 1990, Glimcher and Kara 1992) are also double underlined in Figure 5.9. The sequence of both the sheep and the human *DOB* core S boxes differ from the consensus GGACCTT/C. The sheep sequence is TGACTTT, while the human sequence is CAACTTT. Exons 4-6 of the *HLA-DOB* gene occur within 1.1 kbp of the end of exon 3. We were unable to recognise any comparable open reading frames in the same region in the sheep, nor did the full length *HLA-DOB* cDNA probe (Tonnelle et al. 1985) hybridise to a further 0.7 kbp downstream of the sequenced region reported here (data not shown). Either intron 3 in the sheep gene is unusually long, or the sheep *DOB* gene has suffered a deletion, or thirdly a spurious ligation has occurred during the construction of the cosmid library. The similarity between the second intron of the human *DOB* gene and the sheep gene in cosmid 305 was 61% over its entire length. On the other hand, while the first 190 bp of the third introns from both genes were 66% similar, the remainder of the sheep sequence was not significantly similar to that of the human sequence. The similarity between the sheep and the human intron 3 sequences ended at an Mbo1 site, GATC, at position 1433 in Figure 5.9. As Mbo1 was the enzyme used to partially restrict the genomic DNA used in the cosmid library construction, there was a possibility that there had been a spurious ligation in clone 305 which resulted in the loss of exons 4-6. If this was the case, then the restriction fragments produced by the digestion of genomic and cloned DNA which hybridised to a *DOB* probe would not match. The result of probing a blot of genomic DNA and cosmid DNA both digested with EcoRl, BamHl, Hindlll and Sacl with an *Ovar-DOB* probe containing exon 2 (a 1.6 kbp Pstl fragment, see Figure 5.8), is shown in Figure 5.10. The EcoRl and BamHl fragments were uninformative as they were truncated in the cosmid by the vector. However, the genomic blot clearly contains Hindlll and Sacl fragments of 9 kbp and 5 kbp respectively, which were not present in the cosmid and did not match with the restriction map of the cosmid (Figure 5.7). It would appear therefore that a cloning artefact has occurred in cosmid 305 which has resulted in the loss of exons 4-6 of the gene.

Figure 5.10

Comparison of the restriction fragments from genomic DNA (A), and from cosmid 305 (B) which hybridise to an exon 2-specific *DOB* gene probe. Both blots were washed in 0.2xSSC, 0.1% SDS at 65^0C. The genomic blot (A) was autoradiographed for 3 nights at -70^0C. The clone blot (B) was autoradiographed for 30 minutes at room temperature.

The restriction enzymes used to digest the DNA were 1. EcoRl, 2. BamHl, 3. Hindlll, 4. Sacl. The positions of the λ Hindll molecular weight markers are indicated.



The amino acid translations of exons 1-3 are compared with those of the human, mouse and rabbit *DOB* genes in Figure 5.11 (the translation of the chimpanzee cDNA differed from the human gene at only three positions and is therefore not shown). The amino acid identities between the sheep gene and the *HLA-DOB*, *H-2Ob* and *Orcu-DOB* genes at exons 1-3 are given in Table 5.2 together with a comparison with sheep *DRB* and *DQB* genes.

Table 5.2. The percentage identity between the amino acid sequences of exons 1-3 of the *Ovar-DOB* gene and those of the human (Tonnelle et al. 1985), mouse (Larhammar et al. 1985) and rabbit *DOB* genes (Chouchane et al. 1993). Figures in bold type are the percentage similarities.

	HLA-DOB	*H-2Ob*	*Orcu-DOB*
Exon 1	62 **73**	42 **69**	62 **73**
Exon 2	80 **87**	76 **85**	87 **92**
Exon 3	81 **87**	76 **86**	70 **80**

Figure 5.11

Amino acid translations of the exons of the human, sheep, mouse and rabbit DOB genes. Dashes idicate identity.

* indicates a deletion/insertion and was put in to maintain the alignment.

```
                Signal peptide
HLA-DOB         MGSGWVPWVVALLVNLTRLDSSMTQG
Ovar-DOB        -SPI-------FSSTVL---A-----
H2-OB           --A-RA----------M----F-IE-
Orcu-DOB        -D-K------------I---Y-----


                β1 Domain
                1       10        20        30        40        50        60        70        80        90
HLA-DOB         TDSPEDFVIQAKADCYFTNGTEKVQFVVRFIFNLEEYVRFDSDVGMFVALTKLGQPDAEQWNSRLDLLERSRQAVDGVCRHNYRLGAPFTVGRK
Ovar-DOB        R-------T---------------R------------A-----L-------E--K----L--N-P-I-A---AS--ML--R--K---------R
H2OB            R---N*------------------HLL----------L-----L----------E---D---K------T--A--NM---QK-K-------E-N
Orcu-DOB        *---------------------Q-RV-----------AH----V-------E-------S--N-P-I-----RS--FL--R------------


                β2 Domain
                       100       110       120       130       140       150       160       170       180
HLA-DOB         VQPEVTVYPERTPL*LHQHNLLHCSVTGFYPGDIKIKWFLNGQEERAGVMSTGPIRNGDWTFQTVVMLEMTPELGHVYTCLVDHSSLLSPVSVEW
Ovar-DOB        ----------K--A*-QHR---L-L----------VT--R----Q-E------L--------------P------E--------P--P-------
H2OB            *P------------*-Q-----L----------SV---R------S------LV-------G-T-----I----DI-S---E-PG--R----A-
Orcu-DOB        -P--G--I-----VCGQ-PAA-L-D**-LLS---RVR-LR--------I----L----------TE---------D--------P---------Y
```

There is a putative glycosylation site NGT at amino acid position 20 and the sequence has cysteine residues at positions 15, 79, 117 and 173 which are conserved in all MHC class ll β chains.

5.4 SINE elements close to the *Ovar-DNA* and *-DOB* genes

Cosmid 46.1 contained restriction fragments which hybridised to class ll *B* gene probes indicated by the hatched box in Figure 5.2. The 6.6 kbp EcoRl fragment within this region was cloned into pBS+ and then further subcloned. There were five Sca1 sites within the EcoRl fragment. The Scal fragments were subcloned and sequence data obtained from them and from the EcoRl/Scal fragment at one end of the clone. Database searches showed that this was an area of highly repetitive DNA. The sequences of the three clones from the five Scal sites are compared in Figure 5.12 with the short interspersed nuclear element (SINE) from the bovine proopiomelanocortin gene (btpomc4, Watanabe et al. 1982), which was used as a consensus element designated Bov-A2 by Lenstra et al. (1993). The sheep sequences were highly similar (90-95%) to the Bov-A2 group of sequences, with the 115 bp monomer unit of Watanabe et al. (1982) occurring as a tandem dimer separated by the relatively well conserved spacer motif which conforms to the general sequence $(CACT_n)_3$CATGCATT. In the sequence from subclone 31 there is an extra CACTTT unit. The Scal fragments 31 and 41 have slightly different SINEs at each end of the clone and they are in an inverted orientation relative to one another. Downstream of exon 3 of the *Ovar-DOB* gene in cosmid 305 there is a region of about 400 bp which is similar to a different group of SINE elements, designated Bov-B by Lenstra et al. (1993) and is a contig of the Pstl family of repeats of Majewska et al. (1988) and the art-2 SINE of Duncan et al. (1987). This area was sequenced before it was realised that the *DOB* gene was truncated. The alignment of the sheep and bovine sequences is shown in Figure 5.13. The Bov-B and the Bov-A2 SINEs show high homology only in the 76 bp at the 3' end of the monomer unit (Lenstra et al. (1993).

Figure 5.12

Multiple alignment of the consensus Bov-A2 SINE, with the repeat elements in the sheep Sca1 subclones from the region in cosmid 12 which hybridised to a *B* gene. Dashes indicate identity. The first and second 115 bp monomer units are underlined.

```
                                        ScaI
BTPOMC4   AAGGCAATGGCAACCCACTCCAGTACTCTTGCCTGGCAAATCCCATGGACAGAGGAGCCT
31FOR                          ---------------G-------GC--GTG---------
31REV                          ---------------A------T---A--CA-----G--
41FOR                          -----------T---A------------TG----G----
41REV                          ---------------G-------GC--GTG-------T-
46SCAR    ____________________ ---------------AG---------A--G---------
                                        first monomer unit

BTPOMC4   GGTAGGCTGCAGTCCATGGGGTCGCTAAGAGTCAGACACGACTGAGCGACTTCACTTTCG
31FO      --------------------------------G------CA----------------A
31REV     --------------------------------G--T----------------C-----A
41FOR     ----------------------T----------GC------------------C-----A
41REV     --------------------------------G------------------------A
46SCAR    ___G____A_______C______A________TG_____A___________-------A
          repeats
BTPOMC4   CTTTTCACTTT CA      TGCATTGGAGAAGGAAATGGCAACCCACTCCAGTATTCTT
31FOR     ----------- --      ----C-----------------------------G-----
31REV     ----------T--CTTTCC---G-----------------------------G-----
41FOR     ----------- --      ----------G----------------------G-----
41REV     ----------- --      ----C-----------------------------G-----
46SCAR    ----------- --      ------_____________________________GGCCTC-
                                        second monomer unit

BTPOMC4   GGC TGGAGAATCCCAGGGAT GGGGGAGCCTGGTGGGCTGCCG TCTATGGGGTCGCAC
31FOR     -C-C----------------CC--------------------A ---G-----------
31REV     - CC---------------- C---------------------- -------- -T----
41FOR     -C-C---------T------- ------------------T--- ---------C-A---
41REV     -C- ----------------C ---------------------A ---G-----------
46SCAR    T__C______T_________ CA------------------nnnnG_____A_____A___
                                        repeats
BTPOMC4   CGAGTTGGACACGACTGAAGTGACTT(AGC)6C
31FOR     A----C-------------A--T-- AGTAGC
31REV     A----C---
41FOR     A--------------C----------(AGC)6A
41REV     A----C---
46SCAR    A____A_T_T_T________G_____ --T-T----TCATATAGCATGC
```

Figure 5.13

Alignment of the SINE sequence at the end of the truncated *Ovar-DOB* gene in cosmid 305 with the consensus Bov-B SINE (Lenstra et al 1993). Dashes indicate identity. Dots indicate gaps inserted to maximise the alignment. The underlined area corresponds to the region of homology to the monomer repeat unit.

```
cosmid 305  tagagggaggcgagggcagatgctgacaaggacctcagggtgaatgagaa 1749
Bov-B                                                     ctgc 4

cosmid 305  agaagtcatagcaggagacaggttct.catgaatgtgctgtaaatacacc 1798
Bov-B       --cca-g-a-tt-aa----gct-a--c-t--g-a-gaaa--t-tg--caa 54

cosmid 305  atgacttttttttttgcaattagagagaatgcagtaaatgcagtcttta 1848
Bov-B       cct-ga-agca-a--aaa--gc-----c-t-a-tt-gccaa--.aagg-c 103

cosmid 305  aatatagtcaaagctatggtttttcaatagtcatgtacaggtgtgagag 1898
Bov-B       c--c-------g-----------c--g----------tg-a-------- 153

cosmid 305  ctgggccataaagaacgctgagcactaaag.actgatgctttcaaattgc 1947
Bov-B       t---a-t--------a-------g--g---a-t---------tg--c--t 203

cosmid 305  agtgcaggagaagactcttcagggtcccttgaactgcaagaaaattaaac 1997
Bov-B       g---tt-------------g--a--------g--------g-g--cc--- 253

cosmid 305  cagtcaatccgaaaagaaaccaaccct.gatattcattggaaggactgat 2046
Bov-B       ----------t---g--g-t--gt---g-g------------------- 303

cosmid 305  gctgaagctgaagctccgatactttggtcacttgacatgaagagctgact 2096
Bov-B       ------------a----ag--------c---c---tg------------ 353

cosmid 305  cactggaaaagaccctgatgctgggaaagaccaagggca.aaggagaagg 2145
Bov-B       --t---------------------------ttg------gg--------- 403

cosmid 305  gggcagcagaggatgagatggttggatagcactgccaactcaatggacac 2195
Bov-B       --a-ga---------------------g---tca--g-----------t 453

cosmid 305  caatttgagcaagctctggcagacagggaaggacaggggagcctggcatg 2245
Bov-B       g-g------t--a---ca-g--ttg-t--t--------ag-------g-- 503

cosmid 305  ctgcag                                            2251
Bov-B       ------tccatggggtcgaaagagtcggacatgactgagcgactgaactg 553
```

A BESTFIT of the sheep sequence to the Bov-B sequence revealed 85% identity. A similar comparison with the Bov-A2 sequence showed a 71% identity but only over the 76 bp at the 3' end of the 115 bp monomer unit. The sheep sequence was truncated at the PstI site and so the repeat region has not been sequenced.

5.5 Phylogenetic analysis of the *DNA/DOB* genes.

A phylogenetic tree of the translated second exons from representative human, mouse, cattle and sheep *A* genes, together with the four *DNA* gene sequences is shown in Figure A1.1. The tree was constructed as described in Appendix 1. The (*HLA-DNA* (*H-2Oa, Ovar-DNA*)) cluster occurred in 99.8% of 2000 trials, suggesting that this cluster was distinct from the rest of the tree. The sequence from the tammar wallaby *DNA* gene was placed on a different branch from those of the mammalian *DNA* sequences.

A phylogenetic tree of translated second exons from the four *DOB* sequences together with those from representative sheep and cattle *B* gene sequences is shown in Figure A1.2. The ((*HLA-DOB, H-2Ob)(Ovar-DOB, Orcu-DOB*)) cluster was supported by 99.7% of trials, suggesting that this cluster was distinct from the rest of the tree. Interestingly, the fork which brought together the *DOB* and the *Ovar-DYB* and *Bota-DIB* genes, occurred in 1672 out of 2000 trees (84%).

5.5 Expression of the *Ovar-DNA* and *DOB* genes.

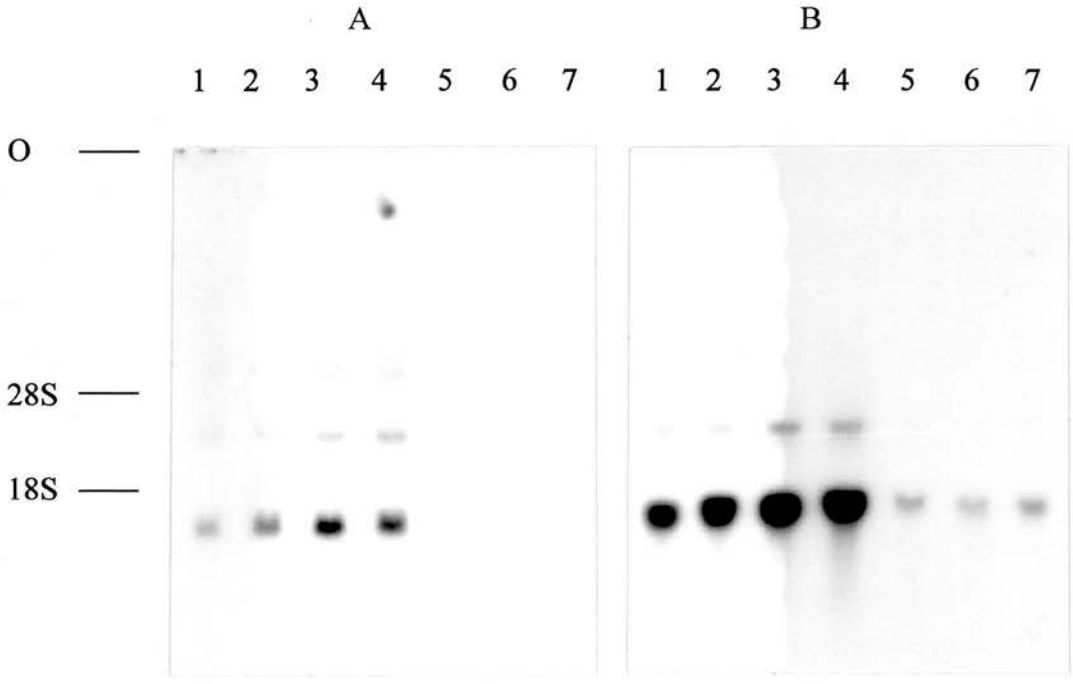
5.5.1 Northern blots.

Total cellular RNA was prepared from sheep thymus, spleen, liver, mesenteric lymph node, kidney, lung and small intestine as described in section 2.3.1. 20 μg of each RNA preparation were run on a denaturing formaldehyde agarose gel and blotted to Hybond-N. The filter was hybridised with a probe which corresponded to the alpha-1

Figure 5.14

Northern blots of total RNA from various sheep tissues hybridised to A, exon 2 probe from the sheep DNA gene, and B, exon 2 probe from the sheep DRA gene. Both blots washed in 0.2xSSC, 0.1% SDS at 65^0C and exposed overnight at -70^0C. 28S and 18S refer to the positions of the two ribosomal RNA bands.

Track no.	Sample	Track no.	Sample
1	Liver	5	Gut
2	Gastric lymph node	6	Lung
3	Thymus	7	Kidney
4	Spleen		



domain of the *Ovar-DNA* gene (the EcoR1/Kpn1 fragment, see Figure 5.3). As a control, a duplicate blot was probed with the *Ovar-DRA* gene. The blots were washed at high stringency (0.2xSSC, 0.1%SDS and 65^0C), and exposed to X-ray film overnight at -70^0C.

The results for the *Ovar-DNA* gene are shown in Figure 5.14. Clearly, *Ovar-DNA* was transcribed but at much lower level than the *Ovar-DRA* gene. The RNA prepared from the thymus, spleen, mesenteric lymph node and liver contained at least an order of magnitude more *DNA* and *DRA* mRNA than that from lung, kidney and gut.

Sizing bands on denaturing gels is difficult without specially prepared standards. However, as indicated in Figure 5.14, the *Ovar-DNA* and *DRA* mRNAs were significantly smaller than the 18S rRNA band which is approximately equivalent to 2 kbp. Both probes also hybridised to a minor band between the 18S and the 28S rRNA bands, (between 2 and 4 kbp).

Using a Pst1 fragment from cosmid 305 which contained exon 2 plus a small piece of exon 3 as a probe, we were unable to detect sheep *DOB* transcripts in Northern blots of RNA from the various tissues inspite of the fact that they were positive for the *Ovar-DRB* isotype mRNA.

5.5.2 Reverse Transcription-Polymerase Chain Reaction (RT-PCR)

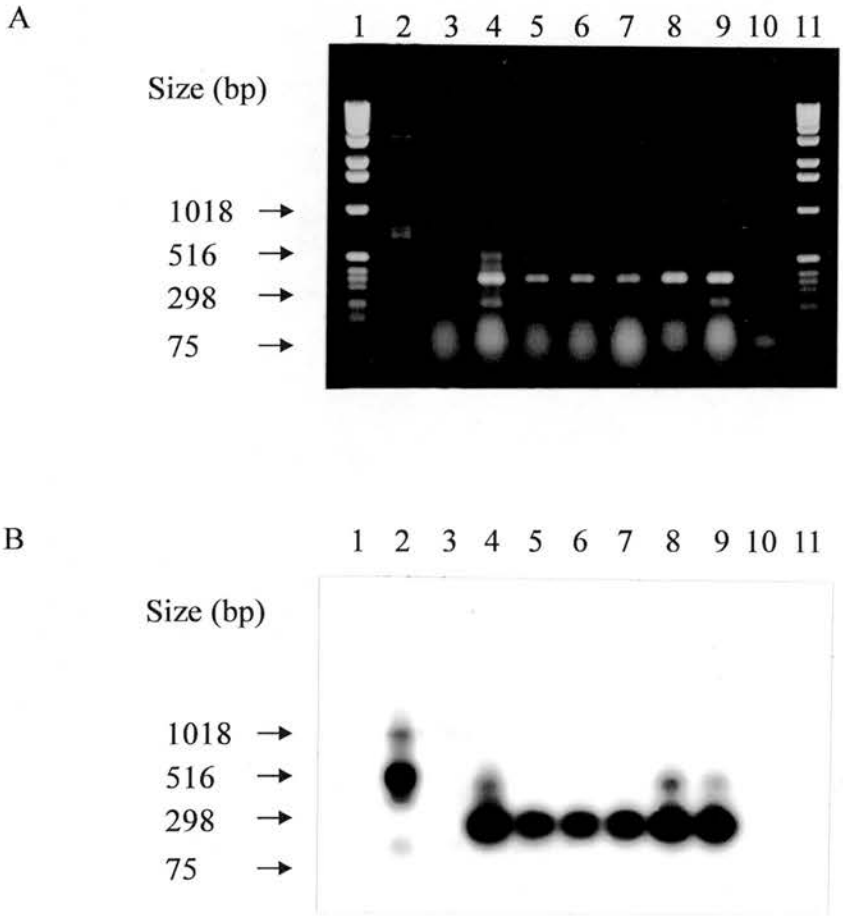
The presence of *Ovar-DNA* mRNA in the total tissue RNA preparations described above was confirmed by RT-PCR. cDNAs were synthesised from 5 μg of total RNA from thymus, spleen, gut, liver kidney, lung and mesenteric lymph node (section 2.3.1). The DNA primers used were DN1 and DN2 whose sequences and orientations are shown in Figure 5.4. The cDNAs were amplified by a standard protocol (section 2.3..2) using these same primers. 1 ng of DNA from cosmid 12, which contains the sheep *DNA* gene was used as a positive control. The expected fragment sizes amplified from the cDNAs and from the gene were 314 and 640 bp respectively, the

Figure 5.15

Evidence for transcription of the sheep MHC class ll *DNA* gene using RT-PCR.

A. Amplified fragments separated on an agarose gel. Primers for both cDNA synthesis and PCR were DN1 and DN2 as described in Figure 5.4. PCR conditions are described in the text.

B. An autoradiograph of gel A following blotting and probing with a sheep *DNA* exon 2 probe from the subclone shown in Figure 5.3. Hybridisation and washing conditions are described in the text.



Track	Sample	Track	Sample	Track	Sample
1	markers	5	gut	9	Thymus
2	cosmid 12	6	liver	10	negative
3	lung	7	kidney	11	markers
4	spleen	8	gastric lymph node		

difference being due to the presence of intron 2 in the gene. The sizes of the fragments obtained are shown in Figure 5.15a and correspond with the expected figures. The track containing the negative control was clear, while the sample tracks, with the exception of that representing lung tissue, contained a band at approximately 300 bp. The band in the cosmid 12 control was slightly larger than the 560 bp standard.

Figure 5.15b shows the result of hybridising the gel in 5.15a with an *Ovar-DNA* specific probe consisting of the Pvull/Kpnl fragment which contained exon 2. Following a high stringency wash, strong signals were obtained in the sample tracks and in the positive control after only 60 minutes exposure to X-ray film.

5.5.3 DNA-mediated gene transfer into mouse L-cells.

Prior to the cloning of the sheep *DOB* gene, a number of attempts were made to express the sheep *DNA* gene in conjunction with the expressed *DRB* gene (chapter 4) without success (Ballingall, personal communication). This may have been due to the isotype mis-match or to the lack of a suitable mab. We were also unable to detect expression of a *DN* protein following co-transfection of cosmid 12 and 305. However, this was not surprising when we realised that the *DOB* gene in cosmid 305 was truncated.

5.6 Discussion

Within the human MHC class ll region, the *DNA* and *DOB* genes form an unusual pairing as they are situated so far apart. The *HLA-DOB* gene is 160 kbp telomeric of the *HLA-DNA* gene and both genes are transcribed in the same direction relative to one another. The mouse orthologues, *H-2Oa* and *Ob*, are similarly orientated and are also well separated at 150 kbp (Karlsson and Peterson 1992). In contrast, the *HLA-DQ1* genes are 10 kbp apart and orientated tail to tail (Campbell and Trowsdale 1993) as are the *Ovar-DQ1* genes (Chapter 3, and Wright and Ballingall 1994). The

sheep *DNA* and *DOB* genes described here have not been physically linked. However, if it is assumed that their transcriptional orientation is the same as in mouse and man, it can be calculated that the minimum distance between them is approximately 45 kbp. Structurally, the sheep genes are very similar to their human and murine counterparts in terms of exon and intron size and the position of conserved features such as glycosylation sites and cysteines involved in intramolecular bonding.

The human, mouse, chimpanzee and rabbit *DN* genes are clearly expressed in some context as cDNA clones have been isolated and sequenced (Trowsdale and Kelly 1985, Tonnelle et al 1985, Servenius et al. 1987, Karlsson and Peterson 1992, Kashahara 1989, Chouchane et al. 1993). Chouchane et al. (1993) detected rabbit *DOB* transcripts on Northern blots of spleen, lymph node and appendix. However, in general the level of transcription of these genes is very low being 2-3% of the level of other isotypes. The human (Tonnelle et al. 1985) and chimpanzee (Kashahara et al. 1989) *DOB* cDNAs were obtained from EBV transformed B cell lines. However the level of *HLA-DOB* RNA was more than thirty times lower than that of the *HLA-DRB* isotype. Stone and Muggli-Cockett (1993) were unable to detect *Bota-DOB* RNA in peripheral blood lymphocytes using the heterologous human *DOB* gene as a probe. This observation for ungulates has now been extended to the sheep using a homologous probe, where *Ovar-DOB* RNA was not detected in a number of tissues by Northern blotting. It is possible that *Ovar-DOB* RNA could be detected by a sensitive assay such RT-PCR. However, this was not tested. Although *Ovar-DNA* transcripts could be detected on Northern blots, the amounts of RNA were very low when compared to that of *Ovar-DRA.*

A feature of the *HLA-DNA* and *H-2Oa* genes is a point mutation in the polyadenylation signal producing the sequence ACTAAA, instead of the eukaryotic consensus, AATAAA. The RNA transcript of the *HLA-DNA* gene was unusually long at 3.5 kbp, and this was ascribed to the altered polyadenylation signal (Trowsdale and Kelly 1985). Subsequently however, human and mouse cDNA

clones of the correct, 1.1 kbp, size were sequenced in which a poly-A tail was added a few base pairs downstream of the ACTAAA signal (Johnsson and Rask 1989, Young and Trowsdale 1990, Karlsson and Peterson 1992). Rosen-Bronson and Long (1991) sequenced a 3.5 kbp cDNA clone derived from the *HLA-DNA* gene and showed that a poly(A) tail had been added after a second mutated polyadenylation signal, ATTAAA, 2.3 kb downstream from the first. The *Ovar-DNA* polyadenylation signal, AGTAAA, carries a different point mutation from that in the human and the mouse. This sequence is approximately one third as active in polyadenylation as the consensus sequence and about twice as active as the ACTAAA form found in the human and mouse (reviewed by Wickens 1990).

There are differences in the regulation of transcription of the genes of the *DN* sub-region. The *HLA-DNA* and *-DOB* genes are not co-ordinately expressed. While the *HLA-DNA* gene, in common with the classical class ll genes of the DP, DQ and DR isotypes, is induced by γ-interferon in fibroblastic cells, the *HLA-DOB* gene is not γ-interferon-inducible (Tonnelle et al 1985) nor is the mouse *H-2Ob* gene (Wake and Flavell 1985). Sequence motifs within the proximal promoter region play crucial roles in the regulation of class ll transcription (reviewed by Benoist and Mathis 1990, Glimcher and Kara, 1992). Voliva et al. (1993) showed by using CAT constructs containing the *HLA-DOB* core S box of seven nucleotides in place of the *HLA-DRA* S box, that the level of transcription from the *HLA-DRA* promoter was drastically reduced in EBV transformed B cell lines and in IFN-γ-induced fibroblastic cells. This was attributed to sequence changes in the core S box. The sequence of the *Ovar-DRA* S box is identical to that of the *HLA-DRA* gene. The equivalent sequence in the sheep *DOB* gene is double-underlined in Figure 5.4. The sheep and human *DOB* S boxes show three and four nucleotide changes from the *DRA* S box respectively. It is therefore possible that a similar mechanism silences both genes. On the other hand, a strict homology to the consensus S box is not essential for expression as the core S box in the *H-2Ob* gene, which is expressed, shows six changes compared to that of the *HLA-DRA* gene. The equivalent region of the expressed rabbit *DOB* gene has not been sequenced.

It was not surprising that we were unable to detect expression of an *Ovar-DN* protein by antibody probing and FACScan analysis following co-transfection of cosmids containing the *Ovar-DNA* and the *Ovar-DOB* genes into mouse L-cells. Cosmids containing the *Ovar-DQ1* genes (Wright and Ballingall, 1994) provided a positive control. The mouse is therefore unusual in expressing a *DN*-like protein at the cell surface. The restricted distribution of the mouse *H-2O* protein, in particular the lack of expression on so-called "professional" antigen-presenting cells, macrophages and dendritic cells, together with the presence of equivalent genes in the diverse species examined, led Karlsson et al. (1991) and Karlsson and Peterson (1992) to conclude that these genes have remained functional throughout evolution and to speculate that their function may be different from that of the other class ll molecules. In this chapter we have shown that another large mammalian family, the ungulates, contains orthologues of the *DN/DO* genes and that they are well conserved in terms of sequence similarity to those of the human, mouse and rabbit. However, we have found no evidence to suggest that these genes are functional.

A phylogenetic analysis of the *B* genes showed that all of the *DOB* genes clustered together. In cattle, the *DOB, DYB,* and *DIB* genes, together with the *DYA,* Tcpb1 and LMP2 genes, form a tightly linked group, llb, separated by a recombination frequency of 0.15-0.2 from the group, lla, which contains the *DQ* and *DR* class ll genes (Andersson 1988, Andersson et al. 1988, Stone and Muggli-Cockett 1993, Shalhevet et al. 1995). It is therefore interesting that the branch which brought together the *DOB* group and the ruminant-specific *DIB* group, although not quite statistically significant, is supported by 84% of trials. A much more rigorous analysis is required before one can speculate about the evolutionary implications of this observation. Certainly, *Ovar-DYA* which is the *A* gene partner for *Ovar-DIB*, does not group with the *DNA* genes, but with the *DQA* genes as shown in Figure A1.1.

The phylogenetic tree showed that, with the exception of the marsupial tammar wallaby gene, the *DNA* genes were linked together on the same branch with bootstrap support of 99%. The marsupial gene had been designated as a *DNA* orthologue by

Slade et al. 1994. The marsupial lineage diverged from the placental lineage 100-150 million years ago (Hope et al. 1990). Slade et al. 1994 argued that as the marsupial/placental split occurred 20-70 million years before the placental radiation, it is likely that orthologous genes from marsupials would be less similar to their placental counterparts than the placental genes are to each other. Furthermore, they considered that the MHC class ll *B* genes isolated from another marsupial, the red-necked wallaby (*Macropus rufogriseus*, Schneider et al. 1991) did not represent new gene families, but rather that the *Maru-DBB* gene was the marsupial orthologue of the placental *DOB* genes. This argument balances the alternative explanation that the class ll families of eutherian and metatherian mammals evolved from different ancestral genes.

It was at first thought that there was a *B* gene adjacent to the *Ovar-DNA* gene in cosmid 46.1 (Deverson et al. 1991). However, sequencing has shown this region of the cosmid to contain highly repetitive DNA related to the bovine SINE family of repeats with no indication of any homology to an MHC class ll gene. I have assumed therefore that the observed hybridisation to class ll B gene probes was artefactual and due to an increased 'stickiness' in this area. The short interspersed nuclear elements make up 1-5% of the genome (Deininger et al. 1986) and are useful markers as they are species specific. Their human and rodent counterparts are the Alu, and B1 and B2 repeats respectively. Repeat-mediated recombination is known to generate genetic diversity. It is interesting to note that the non-polymorphic *HLA-DNA* and *-DOB* genes, to which no function has been assigned, are the only class ll genes in the human MHC not to contain repeat elements (Radley et al. 1994). This led these authors to speculate that this may be due to a specialised selection related to function.

The sequences described here have been assigned EMBL accession numbers : *Ovar-DNA*, z29533, and *Ovar-DOB* : numbers z49879 and z49880.

CHAPTER 6

The *DY* sub-region of the sheep MHC contains an *A/B* gene pair.

6.1 Introduction

The cloning and sequencing of sheep and cattle MHC class ll genes (Muggli-Cockett and Stone 1988, 1989, Groenen et al. 1990, van der Poel et al. 1990, Andersson et al. 1991b, Scott et al. 1991a and b, Ballingall et al. 1992, Sigurdardottir et al. 1991a, 1992, Schwaiger et al 1993a,b,1994, van Oorschot et al. 1994, Wright and Ballingall 1994, Wright et al. 1995), has demonstrated that broadly speaking the ruminant genes are similar to their human counterparts. Of more interest therefore, are loci within the ruminant MHC which differ from those of the *HLA* class ll region.

Four distinguishing features of the ruminant class ll region described to date are firstly, the apparent absence of a *DP*-like isotype; secondly, the variability in the number of *DQ* genes between haplotypes (Andersson and Rask 1988); thirdly, the presence of two groups of class ll genes (Andersson 1988, Andersson et al. 1988) and lastly the presence of class ll genes presumed to be unique to the ruminant (Andersson et al. 1988). The presence of two such genes designated *DYA* and *DYB*, was deduced from RFLP studies of bovine DNA. These genes were shown to segregate together with the *DOB* gene in one region, group llb, separated by a recombination frequency of 0.17 from the region, group lla, which contains the *DQA, DQB, DRB, DRA* and *C4* loci (Andersson et al. 1988). Subsequently, *Bota-DYA* was cloned from a phage library and sequenced (van der Poel et al. 1990). The sequence of part of a similar gene in the goat, obtained by PCR using primers derived from the bovine sequence, has also been reported (Mann et al. 1993). However, there has been no report of the cloning of a *B* gene partner for the *DYA* gene.

A novel bovine class ll *B* gene designated *Bota-DIB* was cloned from a phage library and sequenced by Stone and Muggli-Cockett (1990). This was shown to be a single-copy gene of limited polymorphism, which on the basis of RFLP analysis was probably not *Bota-DYB* but did appear to be distinct from other known bovine class ll genes. The species distribution of this *B* gene was shown to be restricted to Cervidae, Giraffidae and Bovidae (Stone and Muggli-Cockett, 1993). However, it is not known

whether or not any of these novel genes are functional. Expressed human class ll genes usually occur as *A/B* gene pairs situated close to each other on the chromosome. This is also the case with *Bota-DQ* genes (Groenen et al. 1990) and *Ovar-DQ* genes (Deverson et al. 1991, Wright and Ballingall 1994). In this chapter we demonstrate that the sheep MHC contains orthologues of the *Bota-DYA* and *DIB* genes, and that in the sheep these genes are situated very close together.

6.2 Cosmid cloning of the sheep *DYA* and *DIB* genes

The cosmid library constructed from DNA prepared from sheep 3 was screened with *Ovar-DQA, Ovar-DQB, HLA-DQA* and *HLA-DQB* gene probes at low stringency. A cosmid clone, 365, was obtained which hybridised weakly to the *Ovar-DQA* and *DQB* gene probes. Restriction digests of the clone for EcoRl, BamHl, Hindlll, Sacl and Smal are shown in Figure 6.1 and the corresponding restriction maps are shown in Figure 6.2. EcoRl fragments of 2.2 and 1.6 kbp hybridised to the *B* gene probe and an EcoRl fragment of 2.5 kbp hybridised to the *A* gene probe as indicated in the map in Figure 6.2.

6.2.1 The *Ovar-DYA* gene

The 2.5 kbp EcoRl fragment was digested with Sau3A and shotgun cloned into pBS+ cut with BamHl. The colonies were screened with an *HLA-DQA* probe at low stringency and eight positive colonies picked for minipreps. Clone 718 was sequenced from the forward and reverse primers and the sequences compared with those in the database. The only highly significant match was to the *Bota-DYA* gene (van der Poel 1990). The forward sequence started at bp 163 in intron 2 of the bovine sequence, while the reverse sequence started at 893 in intron 3. Inspection of the *Bota-DYA* sequence indicated that exon 2 of the sheep sequence was probably not within the 2.5 kbp EcoRl fragment. It was decided therefore to prepare oligonucleotide primers from the cattle sequence to amplify the fragment containing

Figure 6.1

Restriction fragments produced by digestion of cosmid 365 with EcoRl (E), BamHl (B), Hindlll (H), Sacl (Sa) and Smal (Sm). The positions of the λ Hindlll molecular weight markers (M) are indicated by arrows. The units are kbp.

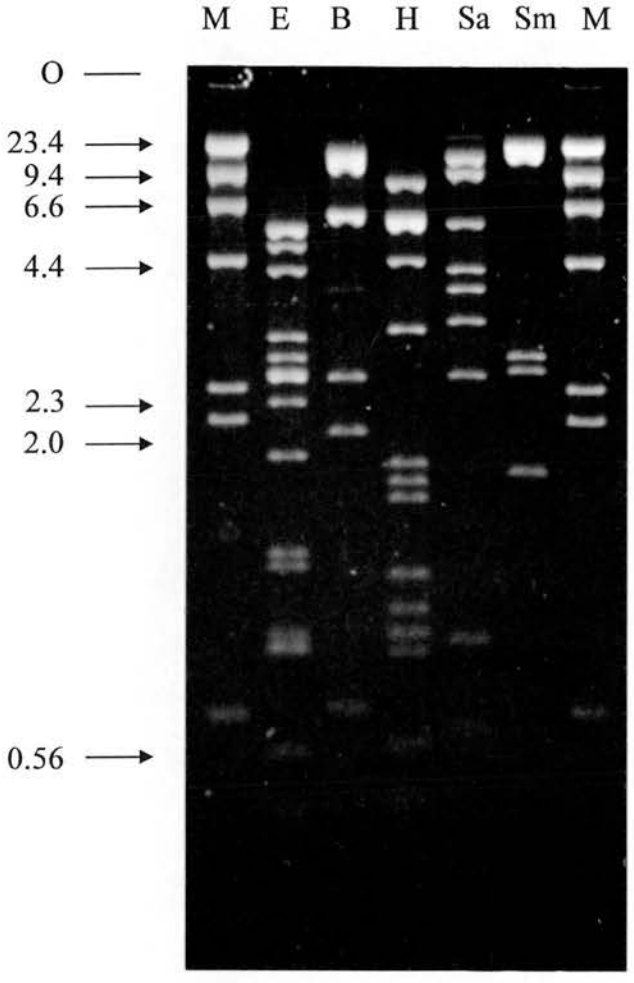


Figure 6.2

Restriction map of cosmid 365 for the enzymes EcoR1, BamH1, Hindlll, Sac1 and Smal. The *A* and *B* genes in cosmid 365 are indicated by unfilled and filled boxes respectively. The likely spatial relationship between the cosmid and the phage clones which contain the *Bota-DYA* gene (van der Poel et al. 1990) and the *Bota-DIB* gene (Muggli-Cockett and Stone, 1990) are also indicated.

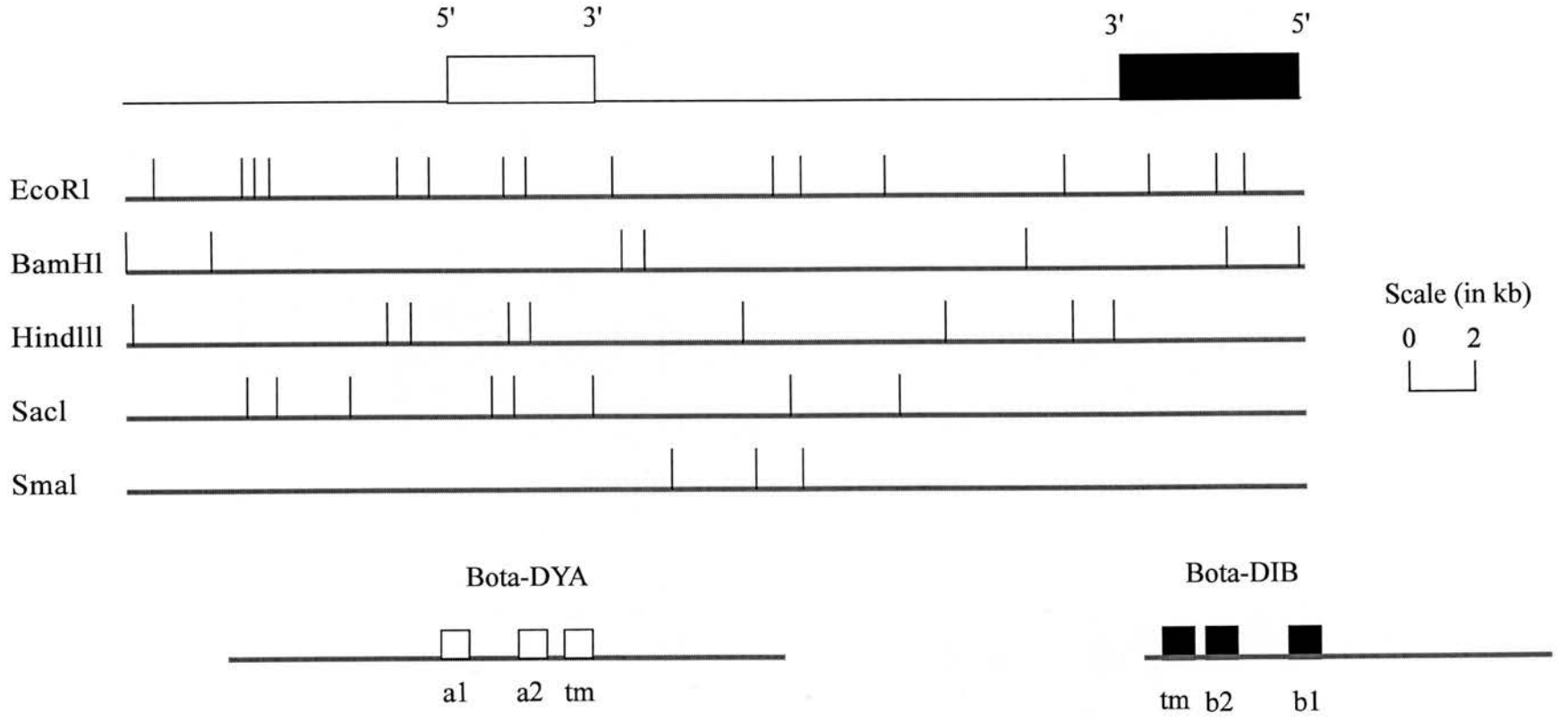


Figure 6.3

Comparison of the nucleotide sequences of exons 2 and 3 of the cattle *DYA* gene (van der Poel et al. 1990) and the *A* gene in cosmid 365. The partial sequence of exon 3 of the goat (Cahi, *Caprinus hircutus*) *DYA* gene (Mann et al. 1993) is also given. Exon sequences are in *capital letters*. Splice junctions are *underlined*. *Dashes* indicate identity. Primers used for sequencing are shown in *bold*. *Italics* indicate that the primer was orientated 3'-5'. Primers DYA1 and DYA2 were also used in RT-PCR.

```
            exon 2
               DYA1
Bota-DYA    agCTGACCACGTGGGCACTTACGGCACAAATGTCTACCAGACGTACGGCA  50
Ovar-DYA    ---G-------------------------------------------

Bota-DYA    CCTCTGGCCAGTTCACGTTTGAATTTGATGGAGACGAGCTCTTCTACGTG 100
Ovar-DYA    ---------G-------------------------------------

Bota-DYA    GACCTGGGGAAAAAGGAGACTGTCTGGCGGCTGTCCGAGTTTAGCAATAT 150
Ovar-DYA    ------A--------------------------C---------A------

Bota-DYA    CACCAAGTTTGAAGTTCAGAGCGCCCTGAGAAACATTGTTATGTCAAAAA 200
Ovar-DYA    ---A-T-------A-------T----------------------------

Bota-DYA    GAAATTTGGACATCTTGATAAAAAATTCCAGCTTTACACCTGCCACCAGT 250
Ovar-DYA    ------------------G----------A----------------AC

Bota-DYA    Ggt    1.19 kb    gaaataatacactaacagctggggagattgtg 300
Ovar-DYA    --- not sequenced --------------g------------g---

Bota-DYA    cacggggtggggaagcctcccctaaactgattaaggagcagaaggcacag 350
Ovar-DYA    --t---------g-----------------g---------t-------g-

                                                    exon 3
Bota-DYA    ttgtcagaaattccaatccatctttggtcttcattgtagAAATCCCTGAA 400
Ovar-DYA    ----------------------------------------C---------

Bota-DYA    GTGGCTGTGTTTCCCAAATCCTCTGTGGTCCTGGGGATTCCCAATACCCT 450
Ovar-DYA    ----------------------C-------------------C-----
Cahi-DYA                                                  --

Bota-DYA    CATCTGTCAAGTGGACAACATCTTTCCTCCTGTGATCAACATCACTTGGT 500
Ovar-DYA    ------------------------------------------------
Cahi-DYA    ------------------------------------------------
```

```
Bota-DYA    TTTACAATGGACACTTTGTTGCAGAAGGGATCGCTGAGACCACCTTCTAC 550
Ovar-DYA    -------------G--------------TG--A-----------------
Cahi-DYA    -------------A--------------TG--------------------

Bota-DYA    CCCAAGAGTGACCACTCTTTCCTCAAGTTCAGTTACCTCACCTTTCTTCC 600
Ovar-DYA    -----------------C---------------------------G----
Cahi-DYA    -----------------C--------------------------------

                                                        DYA2
Bota-DYA    CACCAGTGAAGACTTCTATGACTGCAGAGTGGAGCACTGGGGCCTGGAAG 650
Ovar-DYA    -G-------------------------------------------------

Bota-DYA    AGCCCCTCGTCAAGCACTGGGgtacgtgcattcccaaacccacacccttc 700
Ovar-DYA    G---------------------------------a----a---g------

Bota-DYA    tccacatccaatccacctgcagacagtgtccttccgaatccggcctcc   748
Ovar-DYA    --------------------t------------t------a-----t
```

Figure 6.4

Comparison of the predicted amino acid sequences derived from the nucleotide sequences for exons 2 and 3 of the *Bota-DYA* gene, the *A* gene in cosmid 365, *Ovar-DQA1* and *DQA2* genes (Wright and Ballingall, 1994), the *Ovar-DRA* gene (Ballingall et al. 1992) and the *Ovar-DNA* gene (Wright et al. 1995). The partial sequence of exon 3 from the goat gene is also shown. Dashes indicate identity. * indicates residues not present in the *DR* isotype. Potential glycosylation sites are in bold type.


```
          exon 2
Bota-DYA  DHVGTYGTNVYQTYGTSGQFTFEFDGDELFYVDLGKKETVWRLSEFSNITKFEVQSALRNIVMSKRNLDILIKNSSFTPATS
Ovar-DYA  ---------------A-----------------R--------P--N---M--I-----------------M---N-----N
Ovar-DQA1 --I----V-I-----P--YY-H------E-----E-R------P---KF-S-DP-G----IATV-H--E---QR-NS-A--N
Ovar-DQA2 --F-S---EI--S -P---Y-Q---------------------PM--QFAG-DP-G--SE-ATA-Q-----T-R-N----IN
Ovar-DRA  ---*IIQAEF-LNPEE-AE-M-D-----I-H--MQ--------P--GHFAS--A-G--A-MAVM-A----M--R-NN--N-N
Ovar-DNA  --M-S--PAF--S-DGA----YD---EQ--S---K-R-A----P--G-FAY-DP-NG-VS-A-I-AH-EV-VER-NG-R-PN

          exon 3
Bota-DYA  EIPEVAVFPKSSVVLGIPNTLICQVDNIFPPVINITWFYNGHFVAEGIAETTFYPKSDHSFLKFSYLTFLPTSEDFYDCRVEHWGLEEPLVKHW
Ovar-DYA  D-------------------------------------Q-----VT--------------------V-A---------------G------
Cahi-DYA                        ----------------------Q-----V-----------------------
Ovar-DQA1  V---T--S--P-M--Q------H-------------LR---S-T--VS--S-LI---Y----IN------SDD-V---K------D---L---
Ovar-DQA2  V---T--S--P-M--Q------H-------------LK---A-T--VS--S-L--D------IG------SDD-V---K------D---L---
Ovar-DRA  **---TLL-NKP-E--E------FI-KFS----SV--LR--IP-TD-VSQ-V-L-RD--L-R--H--P----T--V---K------N---L---
Ovar-DNA  VP-R---L---R-Q--E--V---I---------N-T-LH--QP-TQ-VTQ-S--AQP----R--H----V-LAD-----K------DQ--FQ--
```

exon 2 and then clone the fragment into the TA cloning vector pCR making use of the A nucleotide overlap created by the Taq polymerase reaction. The two primers taken from the cattle sequence were D668 (5'-3') and AA040 (3'-5'). The positions of these primers are shown in Figure 6.3. The PCR reaction was as described in section 2.3.3.2, and 1, 10 and 30 pg of denatured cosmid 365 were used as template. The amplified fragment size was about 1.6 kbp in agreement with the size of fragment predicted from the sequence of the cattle gene.

Exons 2 and 3 of the *A* gene in cosmid 365 were then sequenced with the aid of the forward and reverse primers in pCR, pBS+ and purchased oligonucleotides. The sequences are compared with the *Bota-DYA* gene, and a partial sequence from the third exon of the goat (*Cahi-DYA)* gene in Figure 6.3. The predicted amino acid translations of these genes together with those of other published sheep MHC class ll *A* genes are shown in Figure 6.4. The *A* gene in cosmid 365 had many of the salient features of an MHC class ll *A* gene. It has the conserved cysteines at positions 110 and 166 of exon 3 which are involved in disulphide loop formation. There are putative glycosylation sites at positions 79 and 121 which are conserved between isotypes and between species. A possible additional glycosylation site at position 53 is also present in the cattle sequence. Another feature which the sheep gene shares with the *Bota-DYA* gene is the abnormal length of intron 2. When the PCR amplified fragment is corrected for the size of the included exonic sequences, the size of intron 2 comes out at approximately 1.2 kbp. This is about three times greater than the equivalent intron in *HLA-DQA* and *Bota-DQA* (van der Poel et al. 1990).

The *A* gene in cosmid 365 showed a high sequence similarity to the bovine and caprine *DYA* genes and much less so to the *Ovar-DRA* gene (Ballingall et al. 1992), the *Ovar-DQA1* and *DQA2* genes (Scott et al. 1991a) and the *Ovar-DNA* gene (Wright et al. 1995) as shown in Table 6.1. The *A* gene described here is clearly the sheep orthologue of the *Bota-DYA* gene.

Table 6.1. The % identity between exons 2 and 3 of the *Ovar-DYA* gene and those of the *Bota-DYA*, *Cahi-DYA* (goat), *Ovar-DQA1, Ovar-DQA2* and *Ovar-DRA* genes.

	Nucleotides		Protein	
	exon 2	exon 3	exon 2	exon 3
Bota-DYA	95	96	89 (92)	92 (98)
Cahi-DYA	*	98*	*	96*(98)
Ovar-DQA1	75	76	67 (82)	69 (83)
Ovar-DQA2	74	78	66 (76)	71 (85)
Ovar-DRA	68	68	54 (71)	58 (72)
Ovar-DNA			53 (72)	71 (79)

* Only a partial sequence is available for the goat gene. Figures in parenthesis are the % similarity between the amino acid sequences.

6.2.2 The *Ovar-DIB* gene

The 2.2.kbp EcoRl fragment which hybridised to a *B* gene probe was digested with the frequently cutting enzymes Haelll, Pstl and Sau3A. Two Pstl fragments of approximately 700 and 900 bp were produced, both of which hybridised to an exon 3/intron 4 probe. These two fragments were cloned into M13mp18/19. The sequence data obtained indicated that the

smaller fragment contained exon 3 and the larger fragment exon 4 of a sheep gene homologous to the *Bota-DIB* gene described by Stone and Muggli-Cockett (1990).

The 2.2 kbp EcoRl fragment did not appear to contain exon 2 of the *B* gene. Inspection of the restriction map in Figure 6.2 indicated that the *B* gene was very near to end of the cosmid insert. The oligonucleotide primer D159 had been designed from the sequence of the right hand side of the pCos8 vector to allow sequencing back in to the cosmid insert (see section 3.8.1). This primer was therefore used to sequence back into the insert of cosmid 365. As is shown in Figure 6.5, the cosmid insert terminated at a BamHl site 96 bp downstream of the start of exon 2 of the *B* gene in the cosmid.

The sequences of the second, third and fourth exons of the *B* gene in cosmid 365 together with their amino acid translations, are shown in Figure 6.5 together with those of the *Bota-DIB* gene (Stone and Muggli-Cockett 1990). The BamHl site in exon 2 of the sheep gene caused the loss of a part of exon 2, the whole of exon 1 and all the upstream regulatory elements during the cloning procedure. The predicted amino acid translations of exons 2, 3 and 4 are shown together with those of an *Ovar-DQB* (Scott et al. 1991a) and an expressed *Ovar-DRB1* gene (Ballingall et al. 1992) in Figure 6.6. The percent identities between the sheep gene and the *Bota-DIB, Ovar-DQB* and *Ovar-DRB1* genes for each exon at both the nucleotide and amino acid levels are given in Table 6.2. The sequence of the *B* gene in cosmid 365 is clearly very similar to that of *Bota-DIB*, with greater than 96% identity at the nucleotide level at exon 2 and greater than 96% identity at the amino acid level. The nucleotide similarity dropped to 71 and 72% at exon 2 compared to the sheep DQB and DRB genes respectively and this translated to 64 and 68% at the amino acid level.

Figure 6.5

Comparison of the cattle MHC class ll *DIB* gene with the sheep *B* gene in cosmid 365. Nucleotide sequences are given together with the predicted amino acid translation above each codon in the case of the cattle sequence, and below each codon in the case of the sheep sequence. Dashes indicate identity in the nucleotide sequence, while * indicates identity in the amino acid sequence. Dots indicate gaps inserted to maintain the alignment. Exon sequence is in *capital letters*. Splice junctions are *underlined*.Oligonucleotides used for sequencing are in ***bold,*** and *italics* indicates that they are orientated 3'-5'.

```
           exon 2
                N  F  V  Y  Q  F  K  G  M  C  Y  F  T  N  G
Bota-DIB   agAGAATTTCGTGTACCAGTTTAAAGGCATGTGCTACTTCACCAACGGGA  50

           T  E  H  V  R  L  V  A  R  Q  I  Y  N  K  E  E  I
Bota DIB   CAGAGCACGTGAGGCTTGTGGCCAGACAGATCTACAACAAGGAAGAGATC  100
Ovar-DIB                                                 G----
                                                           *

            L  H  F  D  S  D  L  G  E  F  V  A  V  T  E  L  G
Bota-DIB   CTGCACTTTGACAGTGACCTGGGCGAGTTTGTGGCTGTTACAGAGCTGGG  150
Ovar-DIB   --------------C----------------------------------
            *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

              R  V  C  A  E  I  W  N  T  Q  K  D  L  L  A  E
Bota-DIB   CCGGGTGTGTGCGGAGATCTGGAACACCCAGAAGGACCTCCTGGCGGAGT  200
Ovar-DIB   ---------C--T-----------------------------------
              *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

           F  R  A  Y  V  D  T  L  C  R  H  N  Y  K  E  T  A
Bota-DIB   TTCGGGCCTACGTGGACACGCTGTGTAGACACAACTACAAAGAGACGGCC  250
Ovar-DIB   ---------------G----------------------G---------
           *  *  *  *  *  E  *  *  *  *  *  *  *  *  *  *  *

            G  F  T  V  Q  R  R
Bota-DIB   GGCTTCACTGTCCAGCGGAGAGgt.....2.7kb.....  274
Ovar-DIB   ---------------A--------.....................
            *  *  *  *  *  *  *

                                          exon 3
                                          V  E  P  T  V  T
Bota-DIB   cctcattctctaattttgtctcgtttcctcctagTGGAGCCTACAGTGAC  324
Ovar-DIB   -g----------g---------t-------------------------
                                          *  *  *  *  *  *
```

```
                V  S  P  A  S  T  E  A  L  N  H  H  N  L  L  V
Bota-DIB    TGTCTCTCCAGCCAGTACAGAGGCCCTGAACCACCATAATCTGCTGGTCT  374
Ovar-DIB    G-----------------------C-----T----------------
              *  *  *  *  *  *  *  *  P  *  L  *  *  *  *  *

            C  S  V  T  D  F  Y  P  R  Q  V  K  V  K  W  F  R
Bota-DIB    GTTCAGTGACAGATTTCTACCCTCGCCAAGTTAAAGTCAAATGGTTCCGG  424
Ovar-DIB    -----------A-------T--------------A-------------
            *  *  *  *  N  *  *  *  *  *  *  *  I  *  *  *  *

             N  Q  Q  E  Q  T  A  G  V  G  F  T  P  L  T  Q  N
Bota-DIB    AATCAACAGGAGCAGACAGCTGGAGTTGGGTTCACACCTCTTACTCAGAA  474
Ovar-DIB    ------------------------------------------------
             *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

               G  D  W  T  Y  Q  I  H  V  M  L  E  T  V  P  Q
Bota-DIB    TGGGGACTGGACCTACCAGATTCACGTGATGCTAGAGACAGTTCCACAGC  524
Ovar-DIB    -------------------------------------A-A-------
              *  *  *  *  *  *  *  *  *  *  *  *  *  I  *  *

            L  G  D  V  Y  V  C  H  L  D  H  P  S  L  Q  S  P
Bota-DIB    TTGGAGACGTCTACGTTTGCCACCTGGACCACCCCAGCCTCCAGAGCCCC  574
Ovar-DIB    -------------T---------G-------------------------
            *  *  *  *  *  *  *  *  V  *  *  *  *  *  *  *  *

             I  T  V  E  W
Bota-DIB    ATCACAGTAGAATGGCgtaagggccgcttcactgaccctacggacccgac  624
Ovar-DIB    -------------------------------t---.---------t----
             *  *  *  *  *

Bota-DIB    aggaaagaaaaagttcagggagagcgctgggtctggtgggggtggactcc  674
Ovar-DIB    --a---------a----------t---------------------g---

Bota-DIB    gtcttcatccctcgtgtcctatgtaactccctgatacaatttctgggctg  724
Ovar-DIB    -------------a------g------c----------------------

Bota-DIB    gaagtgaccgaggactagatcccagtat..............catctgt  774
Ovar-DIB    ----------------------------ctcaaagttgaaagg------c

Bota-DIB    tgagtccttatctcatttcccccccccaagatgtgatggtggtcccttca  824
Ovar-DIB    ------------------..t------------g---------------

Bota-DIB    cacgaccggacccccccc....not sequenced in cattle....  874
Ovar-DIB    --t--t--------t-t---ctccctccaggctgcagttcagcctgagta

Bota-DIB    ..........................gggctttgacgtagaggctggt.  924
Ovar-DIB    atgctccttctgaagtgttcaagctct----------a----------c

Bota-DIB    atctgaa.gtgacccgtggaggacagacaaacagatggacatgccctctt  974
Ovar-DIB    -------g---------------------...-----------tc---
```

```
Bota-DIB   gg..ggctaccagcct.cccccaccatccatc.gagccccccgg.gcctt 1024
Ovar-DIB   --ga-a--g---t---c-----gt----t----t------t---at-tt-

Bota-DIB   gcctgccctcccctcttctgtctggtttctgatcccctctgtctcttata 1074
Ovar-DIB   ---c--t-------------------------g----------------

Bota-DIB   cacacaacc**agactccaggcttccag**acgaggatgctg..tgggccgtgg 1124
Ovar-DIB   ----tg----------------------a---------ata---------
                                        exon 4
                                        R  A  Q  S  E  S  A
Bota-DIB   ggacactgacctgg..ggcttttacttccagGGGCACAGTCTGAATCAGC 1174
Ovar-DIB   ---ac----gc---ga-------------------------------T-T
                                        *  *  *  *  *  *  V

             Q  S  K  M  Q  S  G  I  G  G  F  V  L  G  L  I
Bota-DIB   CCAGAGCAAGATGCAGAGCGGAATTGGAGGTTTTGTGCTGGGGCTGATCT 1224
Ovar-DIB   -------------TG-----------------------------------
             *  *  *  *  W  *  *  *  *  *  *  *  *  *  *  *

           F  L  G  V  G  L  F  V  H  F  W  D  K  R
Bota-DIB   TCCTTGGTGTGGGCCTTTTTGTCCACTTTTGGGATAAGAGAGgtaaggcg 1274
Ovar-DIB   ----------------------------CA-------------g---a
           *  *  *  *  *  *  *  *  *  *  Q  *  *  *

Bota-DIB   ccttgggagaaaaaggggggaagacaggtgctgggctgaaaagcctctgtt 1324
Ovar-DIB   -------------.-------------c---a-----------------

Bota-DIB   gatccttctctagtgactgtctcagtgacgttggggtgatcttcttttgt 1374
Ovar-DIB   --a--------g-------------------------------------

Bota-DIB   ggcagttgaatccttattgtatgtgggtgggagcaaagaacgctctggaa 1424
Ovar-DIB   ---------------t---------t-----------------------

Bota-DIB   ctgccccctcatt                                     1438
Ovar-DIB   -------------
```

Figure 6.6

The predicted amino acid translations of exons 2-4 of the cattle and sheep *DIB* genes, compared with those of the expressed sheep *DQB1* and *DRB1* genes. Note that the sheep *DIB* gene was truncated during the cloning process.

```
          exon 2
Bota-DIB  NFVYQFKGMCYFTNGTEHVRLVARQIYNKEEILHFDSDLGEFVAVTELGRVCAEIWNTQKDLLAEFRAYVDTLCRHNYKETAGFTVQRR
Ovar-DIB                                  ----------------------------------------E-----------------
Ovar-DQB  D--FL-M-Q--------R----T-Y---Q--H-R----W--YR---PP-QRQ--YF-S---I-ERTG-EA--V-----QVE-A--W---
Ovar-DRB  H-LEYT-KE-R-S----R--FLD-YF--G--YVR----W--YR--A----PD-KY--S--EI-ERR-TE---Y-----GVIES-S----

          exon 3
Bota-DIB  VEPTVTVSPASTEALNHHNLLVCSVTDFYPRQVKVKWFRNQQEQTAGVGFTPLTQNGDWTYQIHVMLETVPQLGDVYVCHLDHPSLQSPITVEW
Ovar-DIB  --------------P-L---------N-------I----------------------------------I----------V-------------
Ovar-DQB  ---------SR-------------------G-I--R----DR-E----VS---IR-----F--L----MT--R----T-RVE--------S---
Ovar-DRB  ---I---Y--K-QP-Q---------NG---GHIE-R----GH-EE---IS-G-I------F-TM--------S-E--T-QV----RT-------

          exon 4
Bota-DIB  RAQSESAQSKMQSGIGGFVLGLIFLGVGLFVHFWDKR
Ovar-DIB  ------V----W--------------------Q---
Ovar-DQB  -----------L--V----------SL--II-HRSQK
Ovar-DRB  --R-D------M--V-------L--A----IY-R-QK
```

Table 6.2. The percentage identity between exons 2, 3 and 4 of the *Ovar-DYB* gene and those of *Bota-DIB, Ovar-DQB* and *Ovar-DRB* at both the nucleotide and amino acid levels.

	Nucleotides			Protein		
	exon 2	exon 3	exon 4	exon 2	exon 3	exon 4
Bota-DIB	96	96	95	98 (100)	94 (98)	92 (92)
Ovar-DQB	71	78	81	49 (64)	73 (83)	66 (86)
Ovar-DRB	71	72	74	53 (68)	62 (75)	68 (84)

Figures in parenthesis are the % similarity between the amino acid sequences.

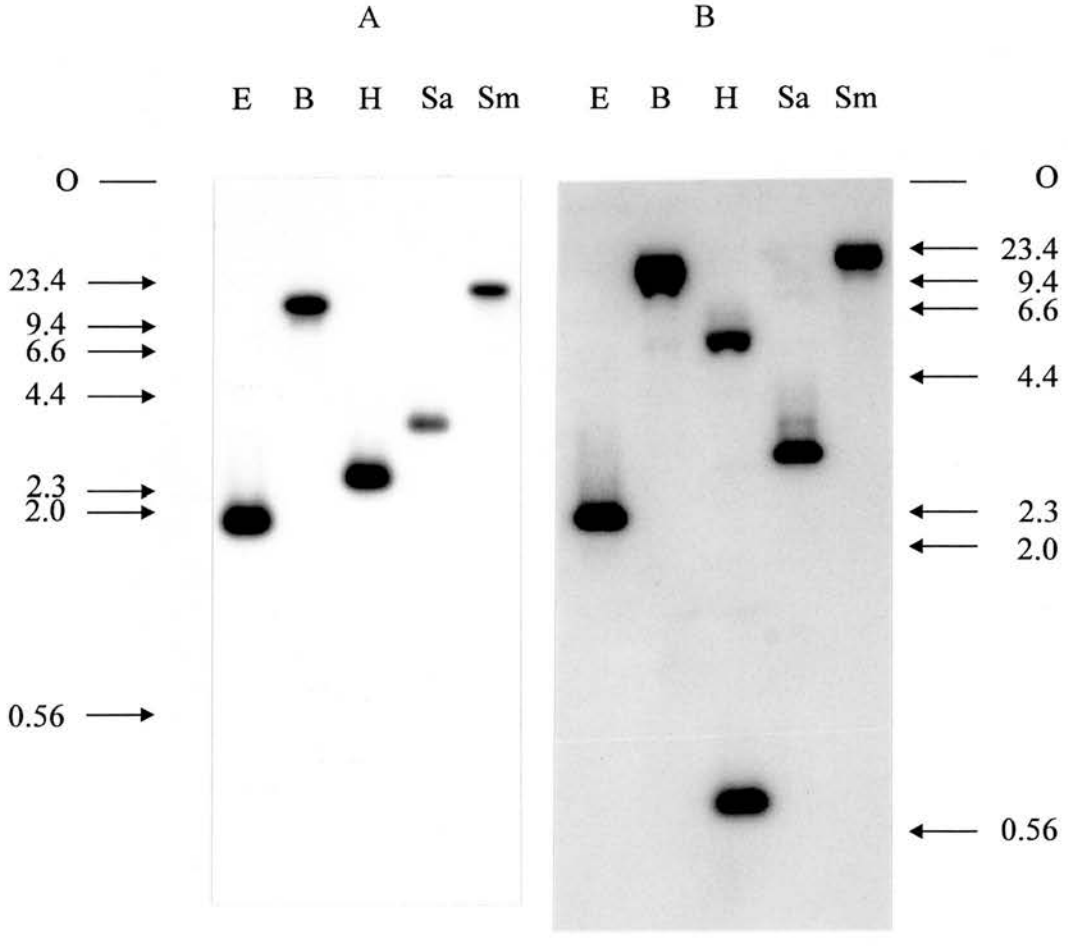
6.3 Transcriptional orientation of the *Ovar-DYA* and *DIB* genes

The sequence of the sheep *DIB* gene in cosmid 365 clearly showed that the gene was truncated by the vector during the cloning process and that the transcriptional orientation was right to left as drawn in Figure 6.2.

The direction of transcription of the *DYA* gene can be determined from Figure 6.7, which shows blots of EcoRl, BamHl, Hindlll, Sacl and Smal digests of cosmid 365 probed with an exon 2-, and an exon 3-specific probe. The Hindlll tracks show that while a 3 kbp fragment hybridised to the exon 2 probe (Figure 6.7a), 6 and 0.6 kbp fragments hybridised to the exon 3 probe (Figure 6.7b). These data, together with the sizes of the other fragments in the same region of the cosmid which hybridise to the *A* gene probes, eg the large BamHl fragment, are only compatible with the

Figure 6.7

Blots of cosmid 365 digested with EcoRl (E), BamHl (B), Hindlll (H), Sacl (Sa) and Smal (Sm) hybridised to A, the *DQA1* exon 2 probe, and B, the *DQA1* exon 3 probe. The blots were washed in 0.2xSSC, 0.1% SDS at 42^0C. The positions of the λ Hindlll molecular weight markers (M) are indicated by arrows. The units are kbp.



arrangement as drawn in Figure 6.2. The *DY* genes are therefore orientated tail to tail and are approximately 11 kbp apart as are the human, sheep and bovine *DQ1* genes.

6.4 Expression of the sheep *DYA/DIB* genes.

6.4.1 DNA-mediated transfection of mouse L-cells

The *Ovar-DR* and *Ovar-DQ* isotypes have been expressed at the surface of mouse L cells following transfection with cosmid DNA (Ballingall et al. 1992, Wright and Ballingall, 1994). The same technique was applied using the *DY* cosmid 365 without success. This was perhaps not surprising when it was realised that the *B* gene was truncated (Figure 6.5). However, since the sheep and cattle genes were clearly very similar as shown above, it was considered possible that the cattle *DIB* gene could participate with the sheep *DYA* gene in the formation of a cross-species heterodimer. The *Bota-DIB* was therefore obtained from Dr R. Stone, USDA, Clay Center, Nebraska, who had obtained the gene on a phage clone in which the gene appeared to be complete. He very kindly sent a sample of a plate lysate from which phage DNA was prepared and used to co-transfect L-cells together with cosmid 365. Results were again disappointingly negative.

6.4.2 Northern blots and Reverse Transcription-Polymerase Chain Reaction

Evidence for the transcription of the *DYA* gene was obtained when total RNA prepared from L cells transiently transfected with the cosmid was hybridised with a *DYA* probe in a slot blot (Ballingall and Wright, 1994). However, transcription of *DYA* could not be detected on Northern blots of RNA from various tissues including thymus and spleen under conditions in which the *DN* and *DR* isotypes were readily detected (chapter 5, section 5.5.1).

To determine whether or not this was simply a question of sensitivity, an RT-PCR assay was set up for *DYA* using the primers DYA1 and DYA2 from exons 2 and 3 the

positions of which are indicated in Figure 6.5. These primers were used for both the cDNA synthesis step and the subsequent PCR step. Under standard PCR conditions, no bands could be detected following blotting and hybridisation to a *DYA* specific probe. Cosmid 365 DNA was used as a positive control and came up as a strong hybridising band. Under similar conditions, *Ovar-DNA* transcription could readily be detected (chapter 5, section 5.5.2).

6.5 Phylogenetic analysis of the *DY/DI* genes

The phylogenetic trees of translated second exons of MHC class ll *A* and *B* genes, constructed as described in Appendix 2, included the sheep and cattle *DYA* and *DIB* genes (Figures A2.1 and A2.2).

Not surprisingly, the *DYA* genes grouped together on a branch which was statistically highly significant, occurring in 2000 out of 2000 trees. As had been reported previously for the cattle gene (Scott et al. 1991a), in general, the *DYA* genes grouped with the *DQA* genes and in this analysis, particularly with the mouse orthologue of the *DQA* genes, *I-A*.

The cluster containing the *DIB* genes was supported by 99.5% of trials confirming that the *DIB* genes are indeed distinct from the rest of the tree. However, while the *DIB* genes generally grouped with the *DQB* genes, their nearest neighbours were in fact those of the *DOB* gene cluster (Figure A2.2).

6.6 Discussion

The MHC class ll genes *Bota-DYA* (van der Poel et al. 1990) and *Bota-DIB* (Stone and Muggli-Cocket 1990), have been described as novel and ruminant-specific. *DYA* has upto now only been sequenced in cattle and in the goat (Mann et al. 1993). The distribution of the *DIB* gene is restricted to Cervidae, Giraffidae and Bovidae (Muggli-Cockett and Stone 1990). Reported polymorphism in the *Bota-DYA* gene is

limited to a G to A transition at position 219 in exon 2, with the ^{219}G allele frequency being 33.3% (Van Eijk et al. 1992). The sheep gene is represented by a G at this position.

The cattle *DYA* and *DIB* genes were isolated on separate phage clones. Cosmid 365 links together the sheep orthologues of these genes and demonstrates that the distance between them is about 11 kbp. The position of the genes in the phage clones relative to those in the cosmid is indicated in Figure 6.2. The arrangement of these genes in cattle could be quite different, although this seems unlikely given the close relationship between the two species. The transcriptional orientation of the genes in the sheep is tail to tail, a structure similar to that of the *HLA-DQ* sub-region (Campbell and Trowsdale 1993), the *Bota-DQ* sub-region (Groenen et al. 1990) and the *Ovar-DQ1* genes (Wright and Ballingall 1994). This may have implications for the evolutionary origin of the *DY/DI* genes. The genes in cosmid 365 show a high degree of sequence similarity to *Bota-DYA* and *Bota-DIB* and were quite different from the *Ovar-DQ, DR, DN* and *DO* isotypes. In the case of *DYA*, this similarity extended to the unusual length of intron 2 (intron 3 is also very long in the cattle gene), and to a possible additional glycosylation site.

There is as yet no estimate of the physical distance between the *DY* locus and the other class II loci in the sheep. Although there is no estimate of the distance between *Bota-DYA* and *Bota-DRB3* in terms of kilobases of DNA, the genetic distance in terms of recombination frequency has been estimated at 0.15 (van Eijk et al. 1993), a figure which agrees well with the 0.17 estimated for the distance between the two sub-regions containing *DY/DO* on the one hand and *DQ/DR* on the other (Andersson et al. 1988). These frequencies contrast with the estimated recombination frequency of only 0.03 between the *HLA-DQ* and *HLA-DP* sub-regions which are separated by 300 kb on the molecular map (Campbell and Trowsdale 1993). Stone and Muggli-Cockett (1993) have shown that the *Bota-DIB* gene maps to the group IIb containing the *DOB* gene and hence, by inference, the *DYA* gene. The evidence presented here is totally in accord with this observation.

Whether an *Ovar-DY* or *Bota-DY* class ll molecule is expressed at the cell surface or not remains to be demonstrated. Given the difference in the primary sequences of the *DY/DI* genes compared to those of the other sheep isotypes, the failure to detect cell surface expression of the sheep *DYA* gene following co-transfection into mouse L-cells with the cattle *DIB* gene, could be ascribed to the lack of a suitable monoclonal antibody. However, a cocktail of mabs was used, including those considered to be pan class ll-specific (SW73.2, Hopkins et al. 1986). It seems unlikely therefore that a class ll molecule would have been missed. Another possibility is that interspecies heterodimer formation is not favoured due to subtle differences in the structures of the α and β protein chains.

Muggli-Cockett and Stone (1993) were unable to demonstrate *Bota-DIB* transcription in bovine PBLs. Similarly, we were unable to demonstrate the presence of sheep *DIB* RNA on Northern blots of various tissues. Furthermore, our inability to detect sheep *DYA* transcription using the sensitive RT-PCR technique argues strongly against expression of sheep *DY* at least in the tissues tested. However, the demonstration of *DYA*-specific RNA in slot blots of total RNA prepared from mouse L-cells following transient transfection with cosmid 365 (Ballingall and Wright, 1994) implied that the sheep gene maybe transcribed in some context.

A phylogenetic analysis of *Bota A* genes by Scott et al. (1991a) placed the *DYA* gene on the same branch as the *DQA* genes but indicated that *DYA* branched off long before the duplication which led to *DQA1* and *DQA2*. The mutations which have accumulated in the *Bota-DYA* gene since it branched off from the *DQ* line led the same authors to conclude that the *DY* genes were unlikely to be functional. On the other hand, sheep and cattle shared a common ancestor 15-20 million years ago (Irwin et al. 1991), and the high degree of sequence similarity between the cattle *DYA* and *DIB* genes and their sheep orthologues provides an argument that like the *DN/DO* genes, these are evolutionary conserved MHC genes of presently unknown function.

In summary, we have shown that the *Ovar-DY* locus is represented by an *A/B* gene pair, a structure similar to that of the *Ovar-DQ1* locus (Wright et al. 1994). In terms of the three possibilities cited by Andersson et al. (1988) to account for the *DY* locus, ie. duplication and transposition of *DQ* or *DR* genes, gene conversion of *DP* genes or lastly, new class ll genes specific to ruminants, the data presented here would favour the duplication of a pair of *DQ* genes.

The sequences described have been assigned EMBL accession numbers; *DYA*, z27398 and z27399, *DYB*, z27400 and z27401. dropped to 71 and 72% at exon 2 compared to the sheep DQB and DRB genes respectively and this translated to 64 and 68% at the amino acid level.

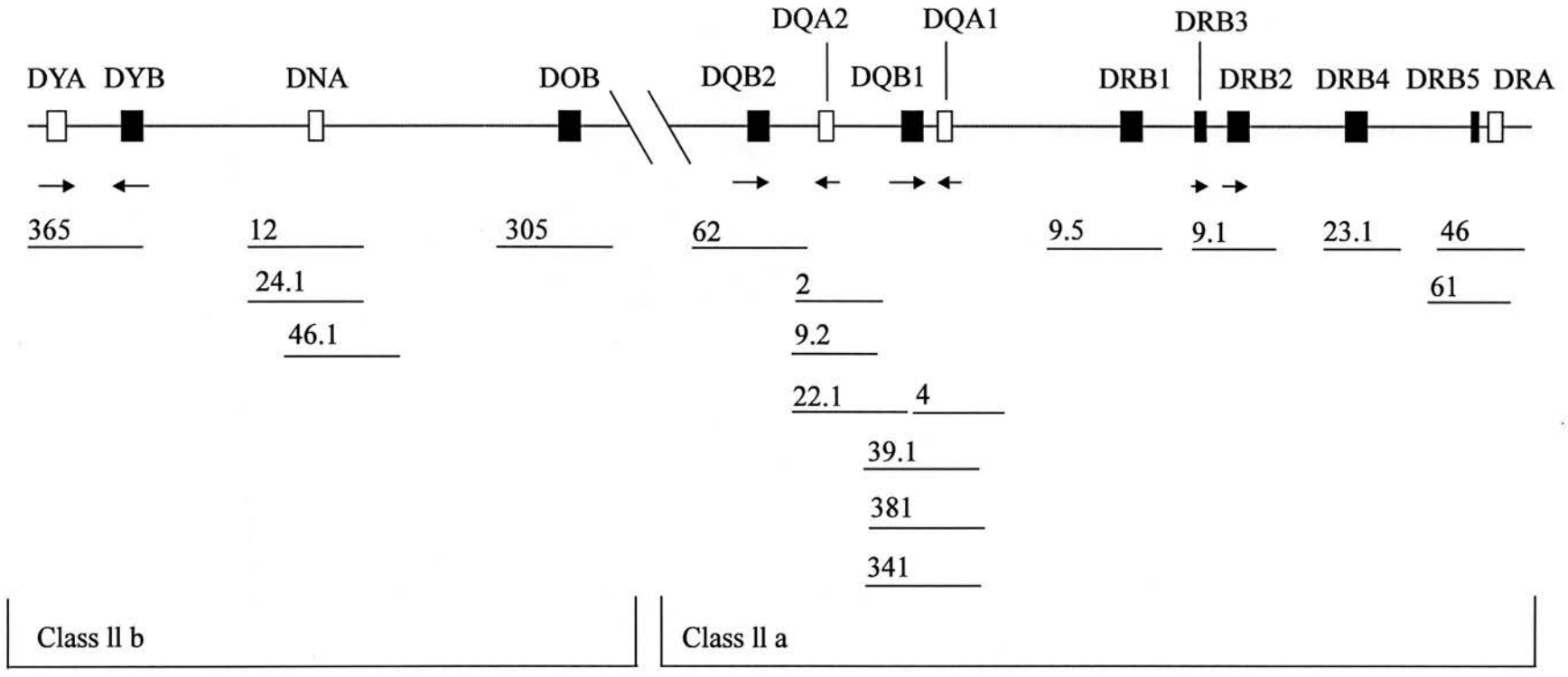
Chapter 7

General Discussion

In this thesis an attempt has been made to define the number and type of MHC class II loci within the sheep MHC by cosmid cloning, DNA sequencing and expression studies. This work has been partly successful as indicated by Figure 7.1 where a map of the sheep class II region is shown. The positions of the cosmid clones which contain the various sheep class II genes which have been described in this work are shown. The map is based on the arrangement of genes in the human MHC class II region as the final arrangement of the genes in the sheep MHC has not been determined. The assumption has been made that the structure of the sheep class II region is similar to that of cattle which contains two groups of class II genes designated a and b (Andersson et al. 1988, Stone and Muggli-Cockett, 1993, van Eijk et al. 1993, Bishop et al. 1994, Shalhevet et al. 1995). This assumption is based on three pieces of information. Firstly, sheep and cattle are relatively closely related in evolutionary terms having last shared a common ancestor 15-20 million years ago (Irwin et al. 1991). Secondly, sheep, cattle and goats contain the ruminant-specific *DY/DI* gene locus which, together with the *DNA* and *DOB* genes, maps to the class IIb group in cattle (Andersson et al. 1988,van der Poel 1990, Stone and Muggli-Cockett 1990, 1993, Mann et al. 1993, Wright et al. 1994). Thirdly, neither sheep nor cattle contain genes equivalent to the *HLA-DP* genes (Scott et al. 1987, Andersson et al. 1988) which map close to the *DNA* gene in man (Campbell and Trowsdale 1993).

The MHC class II genes of sheep and cattle are similar in structure to those of mouse and man. The overall exon/intron structures reflect the extracellular, transmembrane and cytoplasmic domains of the protein and are similar to those in man (Radley et al. 1994). In most cases the conserved features such as glycosylation sites and intra-molecular disulphide bridges occur in the same locations between the various species. All intron/exon boundaries which have been sequenced are of the class I type, that is splicing occurs between the first and second nucleotides of the junctional codon as has been described for other MHC class I and class II genes (Hood et al. 1983). Comparative studies become interesting when they reveal differences which

Figure 7.1



recombination
distance of 0.17

may be related to function. Cattle and sheep both show haplotypic variation in the number and type of *DQ* genes (Andersson and Rask, 1988, Sigurdardottir et al. 1992b, Scott et al. 1991, Fabb et al. 1993). The term haplotype is derived from haploid genotype. Each individual inherits one haplotype from each of its parents. In a highly polymorphic region such as the MHC with many alleles at many loci, there are clearly many different haplotypes. In humans, there is no haplotype variation in the number of *DQ* genes, each haplotype has the full complement of *DQA* and *B* genes (Bohme et al. 1985). Some MHC haplotypes in the sheep on the other hand lack *DQ1* genes (Figure 3.17). Since the *DQ1* genes are functional in the sheep and are the only functional *DQ* genes in the human, this suggests that the *DQ2* genes may also be functional in the sheep. Further indirect evidence for functional sheep *DQ2* genes comes from the fact that the *DQA2* genes in sheep are more polymorphic than the *DQA1* genes (Chapter 3 and Scott et al. 1991a). In man, the *DQ2* genes are not polymorphic and are not functional (Jonsson et al. 1987). There is no evidence that those sheep which lack the *DQ1* locus are at a disadvantage compared with those which have both *DQ* loci. Direct evidence for the expression of the sheep *DQ2* genes resulting in an MHC class ll molecule able to react with the anti-sheep class ll monoclonal antibodies at the surface of peripheral blood monocytess is presented in Chapter 3. It is somewhat surprising therefore that the *DQA2* and *DQB2* genes in cosmids 2 and 62 failed to express following co-transfection into the mouse L cell. It is possible that the alleles present on these two clones are not a favoured pairing and are unable to form a heterodimer at high enough concentrations on the cell surface to be detected. The phenomenon of favoured *A* and *B* allele pairings has been described in the human *DQ* sub-region (Kwok et al. 1993).

In cattle, the structure of the *DQ* sub-region appears to be even more complicated than in the sheep, although this may simply be a reflection of the larger volume of work done on the cattle MHC. Haplotypes have been described which contain one *A* and one *B* gene, two *A* and two *B* genes, and two *A* and one *B* genes. Surprisingly, it appears to be the equivalent of the *DQ2* genes which are absent from those cattle haplotypes which contain only one *A/B* gene pair (Andersson and Rask 1988,

Sigurdardottir et al. 1992). Interestingly, it has been shown that both *DQB* genes are expressed in cattle haplotypes which carry a duplicated *DQ* sub-region (Xu et al. 1994).

MHC class ll genes are co-dominantly expressed, which means that at each of the expressed loci, each *A* allele can form a heterodimer with each *B* allele. The use of DNA-transfected murine L cells to simplify the determination of the specificities of antibodies against the products of complex gene families such as the MHC is well documented and in the case of HLA class ll formed the major part of the 11th International Histocompatibility Workshop (Inoko et al. 1991). The availability of L cell lines which express the *DQ* isotype (Wright and Ballingall 1994) together with the *DR*-specific lines (Ballingall et al. 1992, 1995) has allowed the analysis of the fine specificities of the anti-sheep class ll monoclonal antibodies. Monoclonal antibodies previously thought to be pan-class ll specific (eg SBUII 49.1) did not recognise both *DQ* lines while other *DQ*-specific monoclonal antibodies recognised one allelic *DQ* product but not the other (Table 3.2). The L cell lines have also been useful in demonstrating the expression of two *DRB* genes (Dutia et al. 1994).

The genes of the *DR* sub-region are the most fully characterised of the sheep MHC. The expressed *DRB* locus, designated *DRB1* in the sheep is highly polymorphic with in excess of 40 alleles sequenced at their second exons (Schwaiger et al. 1994). The present study indicates that the sheep *DR* sub-region may contain up to five different *DRB* loci depending upon the haplotype. Only the *DRB1* loci has so far proved capable of expressing a *DR*-protein when co-transfected into the mouse L cell with the sheep *DRA* gene (Ballingall et al. 1992, 1995). The expressed *DRB* locus in cattle is called *DRB3* simply because it was the third locus to be reported (Burke et al. 1991). *DRB3* is also highly polymorphic (Sigurdardottir et al. 1991). Because of the polymorphism of the expressed *DRB* genes, they have been a focus for the development of methods for typing animals at the class ll region. Many of the disease association studies in ruminants have used serologically-determined class l polymorphism as a marker for the MHC as a whole. Serology, RFLP and one

dimensional isoelectric focussing (ID-IEF) of expressed MHC proteins have been compared as typing methods for class II in cattle (Davies et al. 1992). However, a simple PCR-based method has been sought to type sheep and cattle to facilitate studies of the significance of MHC polymorphism in these species (Van Eijk et al. 1992c, Ellegren et al. 1993, Schwaiger et al. 1993). Polymorphism in the length of the microsatellite in the second intron of *DRB* genes together with allele-specific sequence information from the adjacent second exon has been used to develop such methods.

Regardless of the typing method used, the results of the search for an association between MHC and suseptibility or resistance to infectious disease have been disappointing and there are undoubtedly multi-factorial non-MHC effects in the genetic control of disease resistance. A major area of interest in sheep has been the genetic component of resistance to the intestinal parasites *Trichostrongylus colubriformis* and *Haemonchus contortus*. The parasitological parameter commonly measured is faecal egg count (FEC) which is a measure of worm burden. A number of reports described the association between sheep class I serology and FEC, but results were inconsistent (reviewed by Hohenhaus and Outteridge, 1995). The lack of serological reagents for sheep class II typing led to studies at the DNA level by restriction fragment length polymorphism (RFLP). Initially human MHC gene probes were used, and the change was made to homologous sheep probes when they became available. However, Blattman et al. (1993) concluded from a large RFLP study that there was no association between genes in or closely linked to the class II region of the sheep MHC and resistance to *H. contortus* infection in Merino sheep, while Hulme et al. (1991) could find no significant association between sheep class II and FEC resulting from *T. colubriformis* infection. On the other hand the possibilities of the molecular approach are exemplified by the study of persistent lymphocytosis caused by bovine leukaemia virus. Initially, susceptibility or resistance to viral infection was thought to be associated with bovine class I alleles (Lewin et al. 1988). However, the class I and class II genes are in strong linkage disequilibrium and the association was subsequently determined to be with the class II *Bota-DRB2* locus

(Van Eijk et al. 1992b). This has been refined further and the viral infection has now been shown to be associated with mutations at two amino acid residues one of which is in the antigen binding site of the second exon of the *Bota-DRB3* gene (Xu et al. 1993). It may be that associations will only be found with simple pathogens such as viruses which are made up of only a few proteins rather than with complex eukaryotes such as nematodes.

Klein and O'hUigin (1994) questioned the validity of searching for MHC associations with certain types of infectious disease. They identified three types of infectious disease-causing organisms in humans. The first group was comprised of those highly virulent pathogens which they considered became associated with man following the neolithic agricultural revolution. In this group they included *P. falciparum, M. tuberculosis, Yersinia pestis*, smallpox and influenza virus. As it takes on average 1.3 million years to incorporate a non-synonymous substitution (that is one which changes the amino acid sequence) into the peptide binding region of a class I gene and 4.5 million years in the case of a class II gene (Klein et al. 1993) these pathogens are likely to have little if any effect on MHC polymorphism given the time available and hence no MHC associations should be expected. Groups two and three on the other hand, which include pathogens and opportunistic pathogens such as papova virus, *E. coli,* herpes, varicella, have co-existed and co-evolved with their hosts for many millions of years. It seems likely therefore that they have been a major driving force in the evolution of the mammalian immune system including the MHC.

The present study has demonstrated that the sheep MHC contains orthologues of all of the class II isotypes of the human MHC with the exception of the *DP*. Maddox and others (unpublished) have shown that the sheep MHC does contain a *DM* locus. Since the first reports of the analysis of the sheep and cattle MHCs at the DNA level (Scott el al. 1987, Andersson 1988), and despite an increasing number of such analyses, no evidence for ruminant *DP*-like genes has been recorded. The approximate location occupied by the *DP* genes in the human MHC is occupied in

the sheep and cattle MHCs by the *DYA* and *DIB* genes. There is no evidence to suggest that the *DY/DI* genes are related phylogenetically to the *DP* genes. The *DY* genes appear to be an offshoot of the *DQ* group. The *A/B* gene structure in the *DY* group at least in sheep is very similar to that of the expressed *DQ1* genes in both sheep and humans. The lack of expression of *DY* at the transcriptional level in the tissues tested and the lack of success of the attempt to 'force' the expression of sheep *DYA* by transfecting it together with the cattle *DIB* gene does not completely rule out the possibility that a *DY* protein is expressed in some context. Although *HLA-DRA* can substitute functionally for *H-2IEa* in transgenic mice (Lawrance et al. 1989), the association of α and β chains at the cell surface is usually under isotype and allelic constraints. Certain *HLA-DQ1* alleles pair preferentially and are expressed at high levels, while others are expressed only at low levels (Kwok et al. 1993). DNA-mediated transfection of *A* and *B* genes from different species, even within an isotype, may not be a valid procedure.

The *DN/DO* genes like the *DY* genes represent something of an enigma. They are clearly present in all species tested to date. Both the *A* and the *B* gene are monomorphic and the sequences of the genes from the five species now analysed indicate that the genes themselves are highly conserved. However, while the genes from man (*HLA-DNA*, Trowsdale and Kelly 1985, *HLA-DOB* Tonnelle et al. 1985), chimpanzee (*Partr-DOB*, Kasahara et al. 1989), sheep (*Ovar-DNA* Chapter 5), rabbit (*Orcu-DOB*, Chouchane et al. 1993) and rat (*RT1-Oa*, *Ob* Arimura et al. 1995) are transcribed, only in the mouse has a *DN*-protein been identified. However, a definite function has yet to be assigned (Karlsson et al. 1992). The demonstration of a non-classical role for the *HLA-DM* class ll gene product in antigen presentation provides a good example of how important the non-polymorphic, relatively divergent class ll gene loci can be (Morris et al. 1994, Sloan et al. 1995).

An objective of the present study was to produce a detailed map of the sheep class ll region. This has been partially achieved. There are three reasons for this, two of which are related to the scale of the problem. Firstly, the construction of long range

restriction maps by cosmid-walking involves many rounds of screening libraries, restriction mapping of individual cosmids, the preparation of DNA probes from the ends of the cosmid clones which are then used to screen the library to find the next clone to repeat the process. Secondly, it became apparent that like the *HLA-D* region, the sheep MHC class ll region is quite extensive. The *H-2I* region of the mouse is contained within a much smaller tract of DNA. The HLA class ll region extends over 1000 kbp (Campbell and Trowsdale 1993), and the detailed maps which are now available were derived by many groups of workers over a number of years. The two groups of cattle class ll genes represented by *DOB, DNA* and *DY* on the one hand and by *DQ* and *DR* on the other, are separated by a recombination frequency of approximately 0.18 (Andersson et al. 1988, van Eijk et al. 1993, Bishop et al. 1994, Shalhevet et al. 1995). Direct conversion of recombination frequencies into kbp of DNA on a physical map is difficult. However, the *HLA-DP* and *DRB1* genes are separated on the physical map by approximately 700 kbp and on a genetic map by a recombination frequency of 0.08 (Campbell and Trowsdale, 1993, Voltz et al. 1994). Therefore, unless there is one or more recombinational hotspots in the region between *DQ* and *DOB* in cattle and hence by inference in sheep, the distance in terms of kbp between the two groups could be quite large. A third reason has been alluded to in discussion throughout this thesis and relates to the animals from which the DNA was obtained for the construction of the cosmid libraries. Not only were three unrelated animals used, but all three were heterozygous at the MHC to a greater or lesser degree. The problems in matching restriction maps from different cosmids around the same loci would have been circumvented by the use of DNA from an animal which was homozygous at the MHC. The absence of allelic forms would have allowed the immediate identification of different loci. Unfortunately such an animal was not identified prior to the start of this project. On the positive side however, working with heterozygous animals did allow the identification of a number of alleles at some of the polymorphic loci and the identification of haplotypic variation, for example, that around the *DRA* gene where one only haplotype contains *DRB5*. In addition, the expression of different alleles at *DQ1* and *DR* allowed the recognition

of allele-specific monoclonal antibody specificities (Wright and Ballingall 1994, Ballingall et al. 1995).

An obvious solution to the problem of construction of a long range restriction map of the sheep MHC was the use of pulsed field gel electrophoresis (PFGE). The minimum size of restriction fragments which hybridise to both a *DY* and a *DQ* probe or a *DQ* and a *DR* probe would provide an estimate of the distance between these regions. Although they have not been documented in any part of this thesis, a number of attempts were made to map the sheep class ll region using PFGE. There are two distinct stages to PFGE; sample preparation, and the electrophoresis of large DNA fragments. Sample preparation involves embedding cells such as peripheral blood monocytes in plugs of high purity agarose so that the DNA which is purified by subsequent cell lysis and removal of contaminating proteins by digestion with proteinase K is protected from mechanical shearing and remains of very high molecular weight. The DNA can then be digested *in situ* with restriction enzymes such as Notl which have recognition sites which occur infrequently in the genome. The resulting large restriction fragments of up to several hundred kbp can then be separated by field inversion electrophoresis. The electrophoresis equipment used was the BioRad Gene Mapper. Sample preparation and electrophoresis were problem-free as assessed by the smear of genomic DNA fragments obtained following post-run ethidium bromide staining of the gels and by the excellent resolution of DNA molecular weight markers which were either concatamers of the bacteriophage λ genome or the chromosomes from the yeast *Saccharomyces cereviseae* (supplied by BioRad). However, difficulties arose in obtaining signals on an autoradiograph following blotting of the gels and hybridisation to sheep class ll gene probes. It was assumed that the large DNA molecules were not being transferred to the nylon with sufficient efficiency and so a number of different blotting procedures were used but without success and the technique was abandoned. The interpretation of results from such experiments would be greatly facilitated by the use of DNA from homozygous animals and such animals should be identified before the experiments are repeated.

Another technique which became more generally available during the course of the work described here was that of cloning into yeast artificial chromosomes (YACs). These cloning vehicles can accommodate hundreds of kbp of inserted DNA and has been used to clone the HLA-D region (Ragoussis et al. 1991). It will be fascinating to see if another lab takes on the job of cloning the sheep MHC in YACs.

The results of basic cloning, mapping and sequencing studies such as those described here in the sheep and by others in sheep, cattle, pigs and goats and other species have shed light on the genetic organisation and polymorphism of the MHC of an economically important group of animals, the ungulates. Some of this information will be of use in anchoring the genetic maps which are being produced. Such maps will be invaluable in the search for those loci which influence the genetic components of disease resistance and susceptibility. Future work will no doubt determine the final order of most of the MHC genes by a combination of techniques such as YAC cloning, PFGE and FISHing (Fluorescence in situ hybridisation). Indeed the cosmids described here have been sent to colleagues in Germany who intent to use FISH for this purpose. Apart from a detailed map of the class II region itself, to will be interesting to see if the genes which map to the class II region but are involved in class I restricted antigen presentation are similarly arranged in ungulates. It has been shown that the LMP2 proteasome subunit gene maps to the class IIb group in cattle (Shalhevet et al. 1995).

Since they remain an enigma, the possibility that the ruminant-specific *DY* genes and the *DN*/*DO* genes are expressed should be studied more closely in different tissues and in particular in foetal tissue, to determine whether or not they are developmentally regulated and what if any is their function. This would be facilitated by antisera which could be produced to the putative DY and DN proteins by immunising animals with specific synthetic peptides derived from the sequences described here.

References

Aida, Y., Kohda, C., Morooka, A., Nakai, Y., Ogimoto, K., Urao, T., and Asahina, M. (1994) Cloning of cDNAs and the molecular evolution of a bovine MHC class II DRA gene. Biochem Biophys Res Commun. 204, 195-202

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D. (1983) Molecular Biology of the Cell. 2nd Edition. Garland Publishing Inc. New York

Allen, P.M. (1987) Antigen processing at the molecular level. Immunology Today 8, 270-273.

Ammer H, Schwaiger FW, Kammerbauer C, Gomolka M, Arriens A, Lazary S, Epplen JT (1992) Exonic polymorphism vs intronic simple repeat hypervariability in MHC-DRB genes. Immunogenetics 35, 332-340

Andersson L, Bohme J, Rask L, Peterson PA (1986a) Genomic hybridization of bovine class II major histocompatibility genes, 1. Extensive polymorphism of DQ alpha and DQ beta genes. Animal Genetics 17, 95-112

Andersson L, Bohme J, Peterson PA, Rask L (1986b) Genomic hybridization of bovine class II major histocompatibility genes, 2. Polymorphism of DR genes and linkage disequilibrium in the DQ-DR region. Animal Genetics 17, 295-304

Andersson L, Gustafsson, K., Jonsson, A-K., and Rask, L. (1991a) Concerted evolution in a segment of the first domain exon of the polymorphic MHC class ll β loci. Immunogenetics 33, 235-242.

Andersson L, Sigurdardottir S, Borsch C, Gustafsson K (1991b) Evolution of MHC polymorphism, extensive sharing of polymorphic sequence motifs between human and bovine DRB alleles. Immunogenetics 33, 188-193

Andersson L. (1988) Organisation of the bovine MHC class II region as revealed by genomic hybridisations. Animal Genetics 19, 32-34.

Andersson, L and Rask, L. (1988) Characterisation of the MHC class-II region in cattle. The number of DQ genes varies between haplotypes. Immunogenetics, 27, 110-120.

Andersson, G., Larhammar, D., Widmark, E., Servenius, B., Peterson, P.A and Rask, L. (1987) The class II genes of the human major histocompatibility complex, Organisation and evolutionary relationship of the DR-beta genes. J. Biol. Chem. 262, 8784-8758.

Andersson, L., Lunden, A., Sigurdardottir, S., Davies, C.J. and Rask, L., (1988) Linkage relationships in the bovine MHC region. High recombination frequency between class ll subregions. Immunogenetics 27, 273-280

Ando A., Kawai J., Maeda M., Tsuji K., Trowsdale J, and Inoko H. (1989) Mapping and nucleotide sequence of a new HLA class II light chain gene, *DQB3.* Immunogenetics 30, 243-249.

Arimura, Y., Tang, W.R., Koda, T. and Kakinuma, M. (1995) Cloning and analysis of a new rat MHC class ll gene, *RT1.DOα.* Immunogenetics 42, 156-158

Auffray C., Lillie J.W., Korman A.J., Boss J.M Frechin N., Guillemot F., Cooper J., Mulligam R.C and Strominger J.L. (1987) Structure and expression of HLA-DQ alpha and DX alpha genes. Immunogenetics 26, 63-73.

Ayane M, Mengle-Gaw L, McDevitt HO, Benoist C, Mathis D (1986) E alpha u and E beta u chain association, where lies the anomaly? Journal of Immunology 137, 948-951

Bach, F and Hirschorm, K. (1964) Lymphocyte interaction, A potential histocompatibility assay in vitro. Science 143, 813-814.

Bach, F.H., Bach, M.L and Klein, J. (1972) Genetic and immunological complexity of major histocompatibility regions. Science 176, 1024-1027.

Ballingall, K.T. (1991) PhD. Thesis. University of Edinburgh

Ballingall, K.T., H. Wright, J. Redmond, B.M. Dutia, J. Hopkins, J. Lang, E.V. Deverson, J.C. Howard, N. Puri, and D. Haig. 1992. Expression and characterization of ovine major histocompatibility complex class II (OLA-DR) genes. Animal Genetics 23, 347-359.

Ballingall, K.T., Redmond, J. Dutia, B.M., Hopkins, J. and H. Wright. (1995) Analysis of the fine specificities of sheep major histocompatibility complex class ll-specific monoclonal antibodies using mouse L-cell transfectants. Animal Genetics 26, 79-84

Batchelor, J.R and McMichael, A.J. (1987) Progress in understanding HLA and disease associations. British Medical Bulletin 43, 156-183.

Becker J.W., and Reeke G.N. Jr. (1985) Three-dimensional structure of beta 2-microglobulin. Proc Natl Acad Sci U S A. 82, 4225-4229

Benacerraf, B and McDevitt, H.O. (1972) The histocompatibility-linked immune response genes. Science 175, 273-279.

Benacerraf, B. (1978) A hypothesis to relate the specificity of T lymphocytes and the activity of I region-specific Ir genes in macrophages and B lymphocytes. J. Immunol. 120, 1809-1812.

Benoist, C. and Mathis, D. (1990) Regulation of major histocompatibility complex genes , X, Y and other letters of the alphabet. Annu. Rev. Immunol. 8, 681-715

Berdoz J., Tiercy J.M, Rollini P., Mach B., and Gorski J. (1989) Remarkable sequence conservation of the HLA-DQB2 locus (DX beta) within the highly polymorphic DQ subregion of the human MHC. Immunogenetics 29, 241-248

Birnboim, H.C and Doly, J. (1979) A rapid alkaline extraction proceedure for screeening recombinant plasmid DNA. Nucl. Acids Res. 7, 1513-1523.

Bishop M.D. Kappes, S.M. Keele, J.W. Stone, R.T. Sunden, S.L. Hawkins, G.A. Toldo, S.S Fries, R Grosz M.D. Yoo, J et al. (1994) A genetic linkage map for cattle. Genetics 136, 619-639

Bissumbhar B, Nilsson PR, Hensen EJ, Davis WC, Joosten I (1994) Biochemical characterization of bovine MHC DQ allelic variants by one-dimensional isoelectric focusing. Tissue Antigens 44, 100-109

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L and Wiley, D.C. (1987a) Structure of human class-Ihistocompatibility antigen HLA-A2. Nature 329, 506-512.

Bjorkman PJ, and Parham P (1990) Structure, function, and diversity of class I major histocompatibility complex molecules. Annual Review of Biochemistry 59, 253-288

Blattman, AN Hulme, D.J. Kinghorn, B.P. Woolaston, R.R. Gray, G.D. Beh, K.J. (1993) A search for associations between major histocompatibility complex restriction fragment length polymorphism bands and resistance to *Haemonchus contortus* infection in sheep. Animal Genetics 24, 277-282

Bodmer J.G., Marsh S.G., Albert E.D., Bodmer W.F., Dupont B., Erlich H.A., Mach B., Mayr W.R., Parham P., Sasazuki T., et al (1994) Nomenclature for factors of the HLA system, 1994. Tissue Antigens 44, 1-18

Bohme, J., Andersson, M., Andersson, G., Moller, E., Pererson, P.A. and Rask, L., (1985) HLA-DRβ genes vary in number between different DR specificies, whereas the number of DQβ genes is constant. J. Immunology 135, 2149-2155

Braunstein, N.S and Germain, R.N. (1986) The mouse Eβ2 gene, a class ll MHC β gene with limited intraspecies polymorphism and an unusual pattern of transcription. EMBO Journal 5, 2469-2476

Briles, W.E., McGibbon, W.H. and Irwin, M.R. (1950) On multiple alleles affecting cellular antigens in the chicken. Genetics 35, 633

Brookes, C.F and Moore, M. (1988) Differential MHC class-II expression on human peripheral blood monocytes and dendritic cells. Immunology 63, 303-311.

Brown J.H., Jardetzky T., Saper M.A., Samraoui B., Bjorkman P.J., and Wiley D.C. (1988) A hypothetical model of the foreign antigen binding site of class II histocompatibility molecules. Nature 332, 845-850

Brown, J. H., Jardetzky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L. and Wiley, D. C. (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 364, 33-39

Burke, D.T., Carle, G.F. and Olson, M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236, 806-811.

Burke, M.G., Stone, R.T and Muggli-Cockett, N.E. (1991) Nucleotide sequence and Northern analysis of a bovine major histocompatibility complex DRβ-like cDNA. Animal Genetics 22, 343-353

Campbell, R.D. and Trowsdale, J., (1993) A map of the human MHC. Immunology Today 14, 347-352,

Chardon P., Kirszenbaum M., Cullen P.R., Geffrotin C., Auffray C., Strominger J.L., Cohenn D and Vaiman M. (1985) Analysis of sheep MHC using class I, II and C4 cDNA probes. Immunogenetics 22, 349-358.

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J and Rutter, J.W. (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18, 5294-5299.

Cho S.G., Attaya M., and Monaco J.J. (1991a) New class II-like genes in the murine MHC. Nature 353, 573-576

Cho S.G, Attaya M., Brown M.G., and Monaco J.J. (1991b) A cluster of transcribed sequences between the Pb and Ob genes of the murine major histocompatibility complex. Proc Natl Acad Sci USA. 88, 5197-5201

Chouchane L., Brown T., and Kindt T.J., (1993) Structure and expression of a transcribed rabbit class II gene with homology to *HLA-DOB*. Immunogenetics 38, 64-66

Church M.G. and Gilbert, W. (1984). Genomic sequencing. Proceedings of the National Academy of Sciences USA 81, 1991-1995.

Collins T., Korman A., Wake C., Boss J and Kappes D. (1984) Immune interferon activates multiple class II major histocompatibility complex genes and the associated invariant chain gene in human endothelial cells and dermal fibroblasts. Proceedings of the National Academy of Sciences USA 81, 4917-4921.

Collins, J and Hohn, B. (1979) Cosmids, A type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage heads. Proc. Natl. Acad. Sci. USA 75, 4242-4246.

Collins, T., Pober, J.S and Strominger, J.L. (1986) Physiologic regulation of class II major histocompatibility complex gene expression. In HLA class II antigens, a comprehensive review of structure and function. Springer Verlag, Berlin p 14

Cooper DW, van Oorschot RAH, Piper LR, Le Jambre LF (1989) No association between the ovine leucocyte antigen system (OLA) in the Australian merino and susceptibility to Haemonchus contortus infection. Intern. J. Paras. 19,695-697.

Daar A..S, Fuggle S.V., Fabre J.W., Ting A., and Morris P.J. (1984a) The detailed distribution of MHC Class II antigens in normal human organs. Transplantation 38, 293-298

Daar A.S., Fuggle S.V., Fabre J.W., Ting A., and Morris P.J. (1984b) The detailed distribution of HLA-A, B, C antigens in normal human organs. Transplantation 38, 287-292

Das HK. Lawrance SK. and Weissman SM. (1983) Structure and nucleotide sequence of the heavy chain gene of HLA-DR. Proc Natl Acad Sci USA 80, 3543-3547

Dausset, J. (1954) Leuco-agglutinins. IV. Leuco-agglutinins and blood transfusion. Vox Sang 4, 190

Dausset, J. (1958) Iso-leuco-anticorps. Acta Haematologica (Basel) 20, 156

David, C.S., Schreffler, C and Frelinger, J.A. (1973) New lymphocyte antigen system controlled by the Ir region of the mouse H-2 complex. Proc. Natl. Acad. Sci. USA 70, 2509-2514.

Davies, C.J. Andersson, L. Joosten, I. Mariani, P. Gasbarre, L.C. Hensen, E.J (1992) Characterization of bovine MHC class II polymorphism using three typing methods, serology, RFLP and IEF. Eur. J. Immunogen. 19, 253-262

Davis, M.M., Cohen, D.I., Nielson, E.A., Steinmetz, M., Paul, W.E and Hood, L. (1984) Cell-type specific cDNA probes and the murine I region, the localisation and orientation of A-alpha. Proc. Natl. Acad. Sci. USA 81, 2194-2198.

Deininger, P.L. and Daniels, G.R. (1986) The recent evolution of mammalian repetitive DNA elements. Trends in Genetics 2, 76-80

Denaro M, Gustafsson K, Larhammar D, Steinmetz M, Peterson PA, Rask L (1985) Mouse MHC class II gene E beta 2 is closely related to E beta and to HLA-DR beta. Immunogenetics 21, 613-616

Derhaag, PJ. (1990) Genetical factors other than HLA-B27. in "HLA-B27-associated diseases". Scand. J. Rheumatol. Suppl. 87, 122-126

Devereux, J., Haeberli, P and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. Nucl. Acid Res. 12, 387-395.

Deverson, E.V., Wright, H., Watson, S., Ballingall, K.T., Huskisson, N., Diamond, A.G. and Howard, J.C. (1991) Class ll major histocompatibility complex genes of the sheep. Animal Genetics 22, 211-225

De Vries, R.R.P., Schreuder G.M.T., Naipal, A. D'Amaro, J. and van Rood, J.J. (1989) Selection by typhoid and yellow fever epidemics witnessed by the HLA-DR locus. In: Immunobiology of HLA, Vol. 2: Immunogenetics and Histocompatibility, Springer Verlag, New York 461-462

Dikiniene N. and Aida Y. (1995) Cattle cDNA clones encoding MHC class II DQB1 and DQB2 genes. Immunogenetics 42, 75

Duncan, C.H. (1987) Novel Alu-type repeat in artiodactyls. Nucleic Acids Research 15, 1340

Dunham I, Sargent CA, Trowsdale J, Campbell RD. (1987) Molecular mapping of the human major histocompatibility complex by pulsed-field gel electrophoresis. Proceedings of the National Academy of Sciences USA 84, 7237-7241

Dutia B.M., McConnell I., Ballingall K.T., Keating P. and Hopkins J. (1994) Evidence for the expression of two distinct MHC class II *DRB* like molecules in the sheep. Animal Genetics 24, 235-241

Dutia B.M., McConnell I., Bird K., Keating P. and Hopkins J. (1993) Patterns of major histocompatibility complex class II expression on T cell subsets in different immunological compartments 1. Expression on resting T cells. European Journal of Immunology 23, 2882-2888.

Dutia, B.M., Hopkins, J., Allington, M.P., Bujdoso, R and McConnell, I. (1990) Characterisation of monoclonal antibodies specific for alpha and beta chains of sheep MHC class II. Immunology, 70, 27-32.

Ehrlich E., Craig A., Poustka A., Frischauf A.M and Lehrach H. (1987) A family of cosmid vectors with the multy-copy R6K replication origin. Gene 57, 229-237.

Ellegren H, Davies CJ, Andersson L (1993) Strong association between polymorphisms in an intronic microsatellite and in the coding sequence of the BoLA-DRB3 gene, implications for microsatellite stability and PCR-based DRB3 typing. Animal Genetics 24, 269-275

Evans, G.A and Wahl, GM. (1987) Cosmid vectors for genomic walking and rapid plasmid restriction mapping. Methods in Enzymology 152, 604-610.

Fabb, S.A., Maddox, J.F., Gogolin-Ewens, K.J., Baker, L., Wu, M.J., and Brandon, M.R., (1993) Isolation, characterisation and evolution of ovine MHC class ll *DRA* and *DQA* genes. Animal Genetics 24, 249-255

Feinberg, A.P and Vogelstein, B. (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Addendum Anal. Biochem. 137, 266-267.

Feiss M. Fisher RA. Crayton MA. and Egner C. (1977) Packaging of the bacteriophage lambda chromosome: effect of chromosome length. Virology 77, 281-293

Figueroa, F, Gutknecht, J, Tichy, H, and Klein, J. (1990) Class II Mhc genes in rodent evolution. Immunol. Rev. 113, 27-46.

Fink, M.A and Quinn, V.A (1953) Antibody production in inbred strains of mice. J. Immunol. 70, 61-67.

Flavell, R.A., Kooter, J.M., De Boer, E., Little, P.F.R and Williamson, R. (1978) Analysis of the beta-gamma-globulin gene loci in normal and Hb lepore DNA, Direct determination of gene linkage and intergene distance. Cell. 15, 25-41.

Ford, C.H.J. (1974) A serological investigation of sheep leucocyte antigens. J. Immunogenetics 1, 345-354.

Ford, C.H.J. (1975) Genetic studies of sheep leucocyte antigens. J. Immunogenetics 2, 31-40.

Germain R.N and Margulies D.H. (1993) The biochemistry and cell biology of antigen processing and presentation. Annual Review of Immunology 11, 403-450.

Germain, R.N., Bently, D.W and Guill, H. (1985) Influence of allelic polymorphism on the assembly and surface expression of class-II MHC (Ia) molecules. Cell 43, 233-242.

Gibson and Medawar (1943) The fate of skin homografts in man. Journal of Anatomy 77, 299-310

Glimcher, L. and Kara, C.J., (1992) Sequences and factors, A guide to MHC class ll transcription. Annual Review of Immunology 10, 13-49,

Gogolin-Ewens, K.J., Mackay, C.R., Mercer, W.R and Brandon, M.R. (1985) Sheep lymphocyte antigens (OLA). Immunology 56, 717-723.

Goldberg, A.L. and Roche, K.L. (1992) Proteolysis, proteasomes and antigen presentation. Nature 357, 375-379

Gorer, P.A. (1936) The detection of antigenic differences in mouse erythrocytes by the employment of immune sera. Br. J. exp. Path. 17, 42-65

Gorer, P.A. (1937) Further studies on antigenic differences in mouse erythrocytes. Br. J. Exp. Path. 47, 231-250

Gorer, P.A. (1938) The antigenic basis of tumor transplantation. J.Pathol. 47, 231-252

Gorer, P.A., Lyman, S and Snell, G.D. (1948) Studies on the genetic and antigenic basis of tumour transplantation. Linkage between a histocompatibility gene and "fused" gene in mice. Proc. Royal Soc. (series B) 135, 499-405

Graham, F.L and Van Der Eb, A.J. (1973) A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology 52, 456-467

Groenen M.A.M., Van Der Poel J.J., Dijkhof R.J.M, and Giphart M.J. (1990) The nucleotide sequence of bovine MHC class II *DQB* and DRB genes. Immunogenetics 31, 37-44

Gustafsson, K., Widmark, E., Jonsson, A-K. (1987) Class ll genes of the human MHC : evolution of the DP region as deduced from nucleotide sequences of the four genes. J.Biol. Chem. 262, 8778-8786

Hanahan, D. (1984) in 'DNA cloning, A Practical Approach" volume 1. D.M. Glover ed. IRL Press, Oxford

Hardy, D.A., Bell, J.I., Long, E.O., Lindstein, T. and McDevitt, H.O. (1986) Mapping of the class II region of the human histocompatibility complex by pulsed field gel electrophoresis. Nature 323, 453-455.

Hauptfield, V., Klein, D and Klein, J. (1973) Serological identification of an Ir region product. Science 181, 167-169.

Hedges. S.B. (1992) The number of replications needed for accurate estimation of the bootstrap P value in phylogenetic studies. Mol Biol Evol. 9, 366-369

Hediger R., Ansari H.A, and Stranzinger G.F. (1991) Chromosome banding and gene locations support extensive conservation of chromosome structure between cattle and sheep. Cytogenetics and Cellular Genetics 57, 127-134.

Henikoff, S (1987) Unidirectional digestion with exonuclease III in DNA sequence analysis. Methods in Enzymology 155, 156-165

Herrmann B.G and Frischauf A-M. (1987) Isolation of genomic DNA. in Methods in Enzymology 152

Higgins D.Q., Bleasby A.J and Fuchs R. (1992) Clustal-V. Improved software for multiple sequence alignment. Computer applications in the Biosciences 8, 189-191.

Hill, A., Allsopp, CEM., Kwiatkowski, D., Anstey, N.M. et al. (1991) Common West African HLA antigens are associated with protection from severe malaria. Nature 352, 595-600

Hill, AV. Yates, SN. Allsopp, CE. Gupta, S. Gilbert, SC. Lalvani, A. Aidoo, M, Davenport, M. and Plebanski, M (1994) Human leukocyte antigens and natural selection by malaria. Philos.Trans. R. Soc. Lond. B Biol. Sci. 346, 379-385

Hohenhaus, M.A. and Outteridge, P.M. (1995) The immunogenetics of resistance to *Trichostrongylus colubriformis* and *Haemonchus contortus* parasites in sheep. British Veterinary Journal 151, 119-140

Hohn, B and Murray, K. (1978) Packaging recombinant DNA molecules into bacteriophage particles in vitro. Proc. Natl. Acad. Sci. USA 74, 3259-3263.

Hood L., Steinmetz, M., and Malissen, B. (1983) Genes of the major histocompatibility complex of the mouse. Annual Review of Immunology 1, 529-68

Hope R., Cooper, S. and Wainwright, B. (1990) Globin macromolecular sequences in marsupials and monotremes. CSIRO Press. Melbourne. p289

Hopkins, J., Dutia, B.M and McConnell, I. (1986) Monoclonal antibodies to sheep lymphocytes. Identification of MHC Class II molecules on lymphoid tissue and changes in the level of Class II expression on lymph-borne cells following antigen stimualtion in vivo. Immunology 59, 433-438.

Hopkins J., McConnell I., Dalziel R.G., and Dutia, B.M. (1993) Patterns of major histocompatibility complex class II expression by T cell subsets in different immunological compartments. 2. Altered expression and cell function following activation in vivo. Eur J Immunol. 23, 2889-2896

Horn, G.T., Bugawan, T.L., Long, C.M. and Erlich, H.A. (1988) Allelic sequence variation of the HLA-DQ loci: relationship to serology and to IDDM susceptibility. Proc. Natl. Acad. Sci. USA 85, 6012-6016

Houghton AN. Thomson TM. Gross D. Oettgen HF. Old LJ. (1984) Surface antigens of melanoma and melanocytes. Specificity of induction of Ia antigens by human gamma-interferon. J Exp Med. 160, 255-69

Hulme, D.G., Windon, R.G., Nicholas, F.W., and Beh, K.J. (1991) Association between MHC class ll RFLP and *Trichostrongylus* resistance in sheep. In 'Breeding for disease resistance in sheep' Eds. G.D. Gray and R.R.Woolaston. p155-120 Melbourne. Wool Research and Development Corp.

Hyldig-Nielsen JJ, Schenning L, Hammerling U, Widmark E, Heldin E, Lind P, Servenius B, Lund T, Flavell R, Lee JS, Trowsdale J, Schreier PH, Zablitzky F, Larhammar D, Peterson PA, Rask L (1983) The complete nucleotide sequence of the I-E alpha d immune response gene. Nucleic Acids Research 11, 5055-5071

Inoko H., Bodmer J.G., Heyes J.M., Drover S., Trowsdale J. and Marshall W.H. (1991) Joint report on the transfectant/monoclonal antibody component. In Tsuji K., Aizawa., Sasazuki T, eds. *HLA 1991*, Volume 1, Oxford, Oxford University Press, 919-929.

Irwin D.M., Kocher T.D., and Wilson A.C. (1991) Evolution of the cytochrome b gene of mammals. J. Mol Evol. 32, 128-144

Ish-Horowicz, D and Burke, J.F. (1981) Rapid and efficient cosmid cloning. Nucleic Acids Research 9, 2989- 2999.

Jardetzky T.S., Lane W.S., Robinson R.A., Madden D.R., and Wiley D.C. (1991) Identification of self peptides bound to purified HLA-B27. Nature. 353, 326-329

Jones, P.P., Murphy, D., Hewgill, D and McDevitt, H. O. (1979) Detection of a common polypeptide chain in I-A and I-E subregion immunoprecipitates. Mol. Immunol. 16, 51-60.

Jonsson A-K, Hyldig-Nielsen JJ, Servenius B, Larhammar D, Andersson G, Jorgensen F, Peterson PA, Rask L. (1987) Class II genes of the human major histocompatibility complex. Comparisons of the DQ and DX alpha and beta genes. Journal of Biological Chemistry 262, 8767-8777

Jonsson, A-K and Rask L., (1989) Human class ll *DNA* and *DOB* genes display low sequence variability. Immunogenetics 29, 411-413

Kapasi, K. Chui, B. and Inman, RD (1992) HLA-B27/microbial mimicry: an *in vivo* analysis. Immunology 77, 456-461

Kappes, D and Strominger, J.L. (1988) Human class-II major histocompatibility complex genes and proteins. Annual Review of Biochemistry 57, 991-1028.

Karlsson, L. and Peterson, P.A. (1992) The alpha chain gene of *H-2O* has an unexpected location in the Major Histocompatibity Complex. J Exp Med 176, 477-483

Karlsson, L., Suhr, C.D., Sprent, J. and Peterson, P.A., (1991) A novel class ll MHC molecule with unusual tissue distribution. Nature 351, 485-488

Kasahara, M., Klein, D. and Klein, J. (1989) Nucleotide sequence of a chimpanzee *DOB* cDNA clone. Immunogenetics 30, 66-68,

Kaufman J.F., Auffray C., Korman A.J., Shackelford D.A., and Strominger J. (1984) The class II molecules of the human and murine major histocompatibility complex. Cell 36, 1-13

Kearney, J.F., Cooper, M.D., Klein, J., Abney, E.R., Parkhouse, R.M.E and Lawton, A.R. (1977) Ontogeny of Ia and IgD on IgM-bearing B lymphocytes in mice. J. Exp. Med. 146, 297-301.

Kelly AP, Monaco JJ, Cho SG, Trowsdale J (1991) A new human HLA class II-related locus, DM. Nature 353, 571-573

Kimura, M. (1983) "The Neutral Theory of Molecular Evolution" Cambridge University Press, Cambridge.

Klein J (1987) Origin of the major histocompatibility complex polymorphism. The trans-species hypothesis. Hum. Immunol. 19, 155-162.

Klein, J. Satta, Y. O'hUigin, C. and Takahata, N. (1993) The molecular descent of the major histocompatibility complex. Annual Review of Immunology 11, 269-295

Klein J, and O'hUigin C. (1994) MHC polymorphism and parasites. Philosophical Transactions of the Royal Society of London Series B 346, 357-358

Klein J., Bontrop R.E., Dawkins R.L., Erlich H.A., Gyllensten U.B., Heise E.R., Jones P.P., Parnam P., Wakeland E.K and Watkins D.I. (1990) Nomenclature for the major histocompatibility complexes of different species, a proposal. Immunogenetics 31, 217-219.

Klein, J. (1975) Biology of the mouse histocompatibility-2 complex Springer-Verlag

Klein, J. (1979) The major histocompatibility complex of the mouse. Science 203, 516-521.

Klein, J. (1986) Natural history of the major histocompatibility complex. Wiley Interscience publication

Klein, J. and Figueroa, F. (1986) Evolution of the major histocompatibility complex. Crit. Rev. Immunol. 6, 295-386

Konig, R., Huang, L.Y. and Germaine, R.N. (1992) MHC class ll interaction with CD4 mediated by a region analogous to the MHC class l binding site for CD8. Nature 356, 796-798

Korman AJ. Auffray C. Schamboeck A. and Strominger JL. (1982) The amino acid sequence and gene organization of the heavy chain of the HLA-DR antigen: homology to immunoglobulins. Proc Natl Acad Sci USA 79, 6013-6017

Kovats S., Main E.K., Librach C., Stubblebine M., Fisher S.J., DeMars R. (1990) A class I antigen, HLA-G, expressed in human trophoblasts. Science 248, 220-223

Kronenberg, M., Steinmetz, M., Kobori, J. (1983) RNA transcripts for I-J polypeptides are apparently not encoded between the I-A and I-E subregions of the murine major histocompatibility complex. Proc. Natl. Acad. Sci. USA 80, 5704-5708

Kvist, S. Bregegere, F. Rask, L. Cami, B. Garoff, H. Daniel, F. Wiman, K. Larhammar D. Abastado JP. Gachelin G. Peterson PA. Dobberstein B. Kourilsky P. (1981) cDNA clone coding for part of a mouse H-2d major histocompatibility antigen. Proc Natl Acad Sci USA 78, 2772-2776

Kwok W.W., Kovats S., Thurtle P., and Nepom G.T. (1993) HLA-DQ allelic polymorphisms constrain patterns of class II heterodimer formation. J Immunol. 150, 2263-2272

Lahesmaa, R. Skurnik, M. Granfors, K. Mottonen, T. Saario, R. Toivanen, A. Toivanen, P. (1992) Molecular mimicry in the pathogenesis of spondyloarthropathies. A critical appraisal of cross-reactivity between microbial antigens and HLA-B27. Br. J. Rheumatol. 31, 221-229

Lang, J.C., Wilkie, N.M and Spandidos, D.A. (1983) Characterization of eucaryotic transcriptional control signals by assay of Herpes Simplex Virus type 1 thymidine kinase. J. Gen. Virol. 64, 2679-2696.

Larhammar, D., Gustafsson, K., Claesson l., Bill, L., Wiman, K., Schenning, L., Sundelin, J., Widmark, E., Peterson, P.A., and Rask, L. (1982) Alpha chain of HLA-DR transplantation antigens is a member of the same protein family as the immunoglobulins. Cell 30, 153

Larhammar, D., Hammerling, U., Rask, L and Peterson, P.A. (1985) Sequence of gene and cDNA encoding murine major histocompatibility complex Class II gene Aβ2. J. Biol. Chem. 260, 14111-14119.

Lawlor D.A., Zemmour J., Ennis P.D., and Parham P. (1990) Evolution of class-I MHC genes and proteins, from natural selection to thymic selection. Annual Review of Immunology 8, 23-63

Lawrance, S.K. Karlsson, L. Price, J. Quaranta, V. Ron, Y. Sprent, J. Peterson, P.A. (1989) Transgenic HLA-DR alpha faithfully reconstitutes IE-controlled immune functions and induces cross-tolerance to E alpha in E alpha 0 mutant mice. Cell 58, 583-594

Lee J.S., Trowsdale, J and Bodmer, W.F. (1982a) cDNA clones coding for the heavy chain of human HLA-DR antigen. Proc. Natl. Acad. Sci. USA 79, 545-549.

Lee JS, Trowsdale J, Travers PJ, Carey J, Grosveld F, Jenkins J, Bodmer WF (1982b) Sequence of an HLA-DR alpha-chain cDNA clone and intron-exon organization of the corresponding gene. Nature 299, 750-752

Lehrach. H., Diamond. D., Wozney. J.M and Boedtker. H. (1977) RNA molecular weight determination by gel electrophoresis under denaturing conditions, a critical re-examination. Biochemistry 16, 4743-4751.

Lenstra, J.A., van Boxtel, J.A.F., Zwaagstra, K.A. and Schwerin, M., (1993) Short interspersed nuclear element (SINE) of the Bovidae. Animal Genetics 24, 33-39

Levine, B.B., Ojeda, A and Benaceraff, B. (1963) Studies on artificial antigen. The genetic control of the immune response to hapten-poly-L-lysine conjugated in guinea pigs. J. Exp. Med. 118, 953-957.

Lewin, H.A., Wu, M.C., Stewart, J.A. and Nolan, T.J. (1988) Association between BoLA and sub-clinical bovine leukemia virus infection in a herd of Holstein-Freisian cows. Immunogenetics 27, 338-344

Little, C.C and Tyzzer, E.E. (1916) Further experimental studies on the inheritance of sucseptibility to a transplantable tumour carcinoma of the Japanese waltzing mouse. J. Med. Res. 33, 393-435.

Little, C.C. and Johnson, (1922) The inheritance of susceptibility to implants of splenic tissue in mice. 1. Japanese waltzing mice, albinos and their F1 generation hybrids, Proceedings of the Society for Experimental Biology and Medicine. 19, 163-169

Long, E.O., Wake, C.T., Strubin, M., Gross, N., Accolla, R.S., Carrel, S and Mach, B. (1982) Isolation of distinct cDNA clone encoding HLA-DR-beta chains by use of an expression assay. Proc. Natl. Acad. Sci. USA 79, 7465-7469.

Lund, T., Grosveld, F.G and Flavell, R.A. (1982) Isolation of transforming DNA by cosmid rescue. Proc. Natl. Acad. Sci. USA 79, 520-524.

Lundberg, A.S. and McDevitt, H.O. (1992) Evolution of major histocompatibility complex class II allelic diversity, direct descent in mice and humans. Proc. Natl. Acad. Sci. USA 89, 6545-6549

Majewska, K., Szemraj, J., Pluccienniczak, G., Jaworski, J. and Pluccienniczak, A. (1988) A new family of dispersed, highly repetitive sequences in the bovine genome. Bioc. Biop. Acta 949, 119-124.

Malissen, B., Hunkapiller, T and Hood, L. (1983) Nucleotide sequence of a light chain gene of the mouse subregion, Aβ. Science 221, 750-754.

Manley, J.L. (1988) Polyadenylation of mRNA precursors. Biochim. Biophys. Acta. 950, 1-12

Mann, A.J., Abraham, L.J., Cameron, P.U., Robinson, W., Giphart, M.J. and Dawkins, R.L., (1993) The caprine MHC contains *DYA* genes. Immunogenetics 37, 292-295.

Marks, MS. Blum, JS. Cresswell, P (1990) Invariant chain trimers are sequestered in the rough endoplasmic reticulum in the absence of association with HLA class II antigens. J. Cell. Biol. 111, 839-855

Marsh S.G.E and Bodmer J (1993) HLA class II nucleotide sequences, 1992. Immunogenetics 37, 79-94

McDevitt, H.O and Chinitz, A. (1969) Genetic control of the immune response, relationship between immune response and (H-2) type. Science 163, 1207-1208.

McDevitt, H.O., Deak, B.D., Shreffler, D.C., Klein, J., Stimpfling, J.H and Snell, G.D. (1972) Genetic control of the immune response, mapping of the Ir-I locus. J. Exp. Med. 135, 1259-1278.

McNicholas, J., Steimetz, M., Hunkapiller, T., Jones, P and Hood, L. (1982) DNA sequence of the gene encoding the EA Ia polypeptide of the balb/c mouse. Science 218, 1229-1232.

Messing, J. (1983) New M13 vectors for cloning. Methods in Enzymology 101, 20-79

Metzgar, R.S., Bertoglio, J., Anderson J.K., Bonnard, G.D and Ruscetti, F.W. (1979) Detection of HLA-DRw (Ia-like) antigens on human T lymphocytes grown in tissue culture. J.Immunol. 122, 949-953.

Meunier H.F., Carson S., Bodmer W.F., and Trowsdale J. (1986) An isolated beta 1 exon next to the DR alpha gene in the HLA-D region. Immunogenetics 23, 172-180

Millot, P. (1979) Genetic control of lymphocyte antigens in sheep, The OLA complex and two minor loci. Immunogenetics 9, 509-534.

Millot, P. (1984) The OLA major histocompatibility complex of sheep. Expl. Clin. Immunogenetics 1, 31-42.

Monaco J.J. A molecular model of MHC class I restricted antigen processing. (1992) Immunology Today 13, 173-179

Monos, D.S., Czansky, E., Ono, S.J., Radka,, S.F., Kappes, D. and Strominger, J.L. (1995) L cells expressing DQ molecules of the DR3 and DR4 haplotypes, reactivity patterns with mAbs. Immunogenetics 42, 172-180

Morris P. Shaman, J. Attaya, M. Amaya, M. Goodman, S. Bergman, C. Monaco, J.J. and Mellins, E. (1994) An essential role for HLA-DM in antigen presentation by class II major histocompatibility molecules. Nature 368, 551-554

Muggli-Cockett NE, Stone RT (1991) Restriction fragment length polymorphisms in bovine major histocompatibility complex class II beta-chain genes using bovine exon-containing hybridization probes. Animal Genetics 22, 123-136

Muggli-Cockett, N.E and Stone, R.T. (1988) Identification of genetic variation in the bovine major histocompatibility complex DRβ-like genes using sequenced bovine genomic probes. Animal Genetics 19, 213-225.

Muggli-Cockett, N.E. and Stone, R.T., (1989) Partial nucleotide sequence of a bovine major histocompatibility class ll DRβ-like gene. Animal Genetics 20, 361-369.

Murphy DB. Jones PP. Loken MR. and McDevitt HO. (1980) Interaction between I region loci influences the expression of a cell surface Ia antigen. Proc Natl Acad Sci USA 77, 5404-5408,

Nei, M., and T. Gojobori. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418-426.

Neefjes, J.J. Stollorz, V. Peters, P.J. Geuze, H.J. and Ploegh, H.L. (1990) The biosynthetic pathway of MHC class II but not class I molecules intersects the endocytic route. Cell 61, 171-183

Nepom, G.T. and Erlich, H. (1991) MHC class molecules and autoimmunity. Annual Review of Immunology 9, 493-525

Ostergard, H. (1989) Improving genetic disease resistance in farm animals. [edited by van der Zijpp, A. J. and Sybesma, W.]. Current Topics in Veterinary Medicine and Animal Science, 52. 115-123 Dordrecht, Netherlands. Kluwer Academic Publishers.

Orr, H.T., Lopez de Castro, J.A, Lancet, D., Strominger, J.L. (1979) Complete amino acid sequence of a papian soluble HLA, HLA-B7. 2. Sequence determination and search for homologies. Biochemistry 18, 5711-5720

Outteridge, PM., Windon, R.G., and Dineen, J.K. (1985) An association between a lymphocyte antigen in sheep and the response to vaccination against the parasite T. colubriformis. Int. J. Parasitology 15, 121-127

Outteridge, P. Windon, R.G., Dineen, J.K. and Smith, E.F. (1986) The relationship between ovine lymphocyte antigens and faecal egg count of sheep selected for responsiveness to vaccination against T. colubriformis. Int. J. Parasitology 16, 369-374

Outteridge, P. Windon, R.G., and Dineen, J.K. (1988) An ovine lymphocyte antigen marker for acquired resistance to T. colubriformis. Int. J. Parasitology 18, 853-858

Pamer, E.G, Wang, C.R., Flaherty, L., Lindahl, K.F., and Bevan, M.J. (1992) H-2M3 presents a Listeria monocytogenes peptide to cytotoxic T lymphocytes. Cell 70, 215-223

Payne, R. and Rolfs, M.R. (1958) Foetomaternal leucocyte incompatibility. J. Clin. Invest. 37, 1756

Parham, P., Lonen, C.E and Lawlor, D.A. (1988) Nature of polymorphism in HLA-A, B and C molecules. Proc. Natl. Acad. Sci. USA 85, 4005-4009.

Ploegh, H.L., Orr, H.T and Strominger, J.L. (1980) Molecular cloning of a human histocompatibility antigen cDNA fragment. Proc. Natl. Acad. Sci. USA 77, 6081-6085.

Polla B.S., Poljak A., Geier S., Nathenson S.G., Ohara J., Paul W.E., and Glimcher L.H. (1986) Regulation of class II gene expression in a murine pre-B cell line. Fed Proc. 45, 2995-2999

Potter, T.A., Rajan, T.V., Dick, R.F. and Bluestone, J.A. (1989) Substitution at residue 227 of H-2 class I molecules abrogates recognition by CD8-dependent, but not CD8-independent, cytotoxic T lymphocytes. Nature 337, 73-75

Puri, N.K and Brandon, M.R. (1987) Sheep MHC class II molecules. Immunology 62, 575-562.

Puri, N.K., Gogolin-Ewens, K.J and Brandon, M.R. (1987a) Monoclonal antibodies to sheep MHC Class I and II molecules, biochemical characterization of three class I gene products and four distinct subpopulations of class II molecules. Vet. Immunol. Immunopathol. 15, 59-86.

Puri, N.K., Gorrel, M.D and Brandon, M.R. (1987b) Sheep MHC class II molecules, immunochemical characterization. Immunology 62, 567-573.

Puri, N.K., Mackay, C.R and Brandon, M.R. (1985) Sheep lymphocyte antigens (OLA). Immunology 56, 725-732.

Puri, N.K., Walker, I.D and Brandon, M.R. (1987c) N-Terminal amino acid sequence analyses of the A and B polypeptides from four distinct subsets of sheep major histocompatibility complex class II molecules. J. Immunol. 139, 2996-3002.

Rackwitz, R-H., Zehetner, G. Frishauf, A-M. and Lehrach, H. (1984) Rapid restriction mapping of DNA cloned in λ phage vectors. Gene 39, 195-200

Rackwitz, R-H., Zehetner, G., Murialdo, H., Delius, H., Chai, J.H., Poustka, A., Frishauf, A-M. and Lehrach, H. (1985) Analysis of cosmids using linearisation by phage λ terminase. Gene 40, 259-266

Radley E, Alderton RP, Kelly A, Trowsdale J, Beck S. (1994) Genomic organization of HLA-DMA and HLA-DMB. Comparison of the gene organization of all six class II families in the human major histocompatibility complex. Journal of Biological Chemistry 269, 18834-18838

Ragoussis J, Monaco A, Mockridge I, Kendall E, Campbell RD, Trowsdale J (1991) Cloning of the HLA class II region in yeast artificial chromosomes. Proceedings of the National Academy of Sciences USA 88, 3753-3757

Riberdy, JM. Newcomb, JR. Surman, MJ. Barbosa, JA. Cresswell, P. (1992) HLA-DR molecules from an antigen-processing mutant cell line are associated with invariant chain peptides. Nature. 360, 474-477

Ries, O, Kammerbauer C, Roewer L, Steimle V, Andreas A, Albert E, Nagai T, and Epplen, JT. (1990) Hypervariability of intronic simple (gt)n(ga)m repeats in HLA-DRB genes. Immunogenetics 32, 110-116

Rollini, P. Mach, B. Gorski, J. (1985) Linkage map of three HLA-DR beta-chain genes: evidence for a recent duplication event. Proc. Natl. Acad. Sci. USA. 82, 7197-7201

Rosen-Bronson, S. and Long, E.O., (1991) An unusual form of alternative splicing in the HLA-DNA gene. Immunogenetics 33, 124-131

Saitou, N. and Nei, M. (1987) The neighbour-joining method, a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406-425

Salter, R.D., Benjamin, R.J.,Wesley, P.K., Buxton, S.E., Garret, T.P.J., Clayberger, C., Krensky, A.M., Norment, A.M., Littman, D.R. and Parham, P. (!990) A binding site for the T cell receptor CD8 on the α3 domain of HLA-A2. Nature 345, 41-46

Sambrook, J., Fritsch, E.F and Maniatis, T. (1989) Molecular cloning. Cold Spring Harbour, Laboratory Press.

Sang, J.H and Sobey, W.R. (1954) The genetic control of the response to antigenic stimuli. J. Immunol. 72, 52-65.

Sanger F., Nicklen S and Coulson A.R. (1977) DNA sequencing with the chain terminating inhibitors. Proc Natl Acad Sci USA 74, 5463-5467.

Schamboeck A. Korman AJ. Kamb A and Strominger JL (1983) Organization of the transcriptional unit of a human class II histocompatibility antigen: HLA-DR heavy chain. Nucleic Acids Res 11, 8663-8675

Schlessinger, D. (1990) Yeast artificial chromosomes, tools for mapping and analysis of complex genomes. Trends in Genetics 248, 255-258

Schneider S, Vincek V, Tichy H, Figueroa F, Klein J (1991) MHC class II genes of a marsupial, the red-necked wallaby (Macropus rufogriseus), identification of new gene families. Molecular Biology and Evolution 8, 753-766

Schwaiger F-W, Weyers E, Buitkamp J, Ede AJ, Crawford A, Epplen JT (1994a) Interdependent MHC-DRB exon-plus-intron evolution in artiodactyls. Mol. Biol. Evol. 11, 239-249

Schwaiger, F.-W., Weyers, E., Epplen, C., Ruff, G., Crawford, A. and J.T. Epplen. (1993a) On the paradox of MHC-DRB evolution, α-helix and β-sheet encoding regions diverge while hypervariable intronic simple repeats co-evolve with β-sheet codons. J. Mol. Evol. 37, 260-272

Schwaiger, F.-W., J. Buitkamp, E. Weyers, and J.T. Epplen. (1993b) Typing of artiodactyl MHC-DRB genes with the help of intronic simple repeated sequences. Mol. Ecol. 2, 55-59.

Schwartz, R.H. (1985) T-lymphocyte recognition of antigen in association with gene products of the MHC Annual Review of Immunology 3, 237-261

Schwartz, D.C and Cantor, C.R. (1984) Separation of yeast chromosome sized DNA's by pulsed field gradient electrophoresis. Cell 37, 67-75.

Scott P.C., Gogolin-Ewens K.J., Adams T.E. and Brandon M.R. (1991a) Nucleotide sequence, polymorphism, and evolution of ovine MHC class ll *DQA* genes. Immunogenetics 34, 69-79.

Scott P.C., Maddox J.F., Gogolin-Ewens K.J and Brandon M.R., (1991b) The nucleotide sequence and evolution of ovine MHC class ll *B* genes, *DQB* and *DRB*. Immunogenetics 34, 80-87.

Scott, P.C., Choi, C and Brandon, M.R. (1987) Genetic organisation of the ovine MHC class II region. Immunogenetics 25, 116-122.

Servenius, B., Rask, L and Peterson, P.A. (1987) Class-II genes of the human major histocompatibility complex. The DO-beta gene is a divergent member of the class-II beta gene family. J. Biol. Chem. 262, 8759-8766.

Shalhevet, D., Yang Da, Beever, J.E., van Eijk, M.J.T., Runlin Ma, Lewin, H.A. and Gaskins, H.R. (1995) Genetic mapping of the LMP2 proteasome subunit gene to the BoLa class llb region. Immunogenetics 41, 44-46

Shawar S.M., Vyas J.M., Rodgers J.R., Rich R.R. (1994) Antigen presentation by major histocompatibility complex class I B molecules. Annu Rev Immunol. 12, 839-880

Shewey LM, Beaty JS, Andersen LC, Nepom GT. (1992) Differential expression of related HLA class II DQ beta genes caused by nucleotide variation in transcriptional regulatory elements. Journal of Immunology 148, 1265-1273

Short, J.M., Fernandez, J.M., Sorge, J.A. and Huse, W.D. (1988) λ zap, bacteriophage λ expression vector with in vivo excision properties. Nucleic Acids Research 16, 7583-7600.

Shoshani J. (1986) Mammalian phylogeny, comparison of morphological and molecular results. Mol Biol Evol. 3, 222-242

Sigurdardottir S, Borsch C, Gustafsson K, Andersson L. (1992a) Exon encoding the antigen-binding site of MHC class II beta-chains is divided into two subregions with different evolutionary histories. Journal of Immunology 148, 968-973

Sigurdardottir, S., Borsch, C., Gustafsson, K. and Andersson, L. (1991a) Cloning and sequencing of 14 *DRB3* alleles of the bovine MHC by using the polymerase chain reaction. Animal Genetics 22, 199-210.

Sigurdardottir, S., Borsch, C., Gustafsson, K. and Andersson, L. (1992b) Gene duplications and sequence polymorphism of bovine class ll *DQB* genes. Immunogenetics 35, 205-213.

Sigurdardottir, S., Lunden, A and Andersson, L. (1988) Restriction fragment length polymorphism of DQ and DR class-II genes of the bovine major histocompatibility complex. Animal Genetics 19, 133-150.

Sigurdardottir, S., Mariani, P., Groenen, M.A.M., Van der Poel, J.J., and Andersson, L. (1991b) Organisation and polymorphism of bovine major histocompatibility complex class ll genes as revealed by genomic hybridisations with bovine probes. Animal Genetics 22, 465-476.

Slade, R.W., Hale, P.T., Francis, D.I., Marshall Graves, J.A., and Sturm, R.A. (1994) The Marsupial MHC, The Tammar Wallaby, *Macropus eugenii*, contains an expressed *DNA*-like gene on chromosome 1. J. Mol. Evol. 38, 496-505.

Sloan, V.S., Cameron, P., Porter, G., Gammon, M., Amaya, M., Mellins, E. And Zaller, D.M. (1995) Mediation by HLA-DM of dissociation of peptides from HLA-DR. Nature 375, 802-806

Smith, G.E. and Summers, M.D. (1980) The bi-directional transfer of DNA and RNA to nitrocellulose or diazobenzyloxymethyl-paper. Analytical Biochemistry 109, 123-129

Snell, G.D., Cloudman, A.M and Woodworth, E. (1948) Tumour immunity in mice induced with lyophilised tissue as influenced by tumour strain, host strain, source of tissue and dosage. Cancer Res. 8, 429-437.

Spandidos, D.A and Wilkie, N.M. (1983) Host-specificities of papillomavirus, Moloney murine sarcoma virus and simiam virus 40 enhancer sequences. EMBO J. 2, 1193-1199.

Steinmetz M. and Hood L. (1983) Genes of the major histocompatibility complex in mouse and man. Science 222, 727-733

Steinmetz M, Minard K, Horvath S, McNicholas J, Srelinger J, Wake C, Long E, Mach B, and Hood L. (1982a) A molecular map of the immune response region from the major histocompatibility complex of the mouse. Nature 300, 35-42

Steinmetz, M., Winoto, A., Minard, K and Hood, L. (1982b) Clusters of genes encoding mouse transplantation antigens. Cell 28, 489-498

Steinmetz M, Stephan D, Fischer Lindahl K. (1986) Gene organization and recombinational hotspots in the murine major histocompatibility complex. Cell 44, 895-904

Stone, R.T. and Muggli-Cockett, N.E. (1990) Partial nucleotide sequence of a novel bovine major histocompatibility complex class ll beta chain gene, *BoLA DIB*. Animal Genetics 21, 352-360

Stone, R.T. and Muggli-Cockett, N.E. (1993) *BoLA-DIB*, species distribution, linkage with *DOB* and Northern analysis. Animal Genetics 24, 41-45

Swarbrick PA, Schwaiger FW, Epplen JT, Buchan GS, Griffin JF, Crawford AM (1995) Cloning and sequencing of expressed DRB genes of the red deer (Cervus elaphus) MHC. Immunogenetics 42, 1-9

Szybalska, E.H and Szybalski, W. (1962) Genetics of human cell lines IV. DNA-mediated heritable transformation of a biochemical trait. Proc. Natl. Anad. Sci. USA 48, 2026-2034.

Tonnelle, C., Demars, R. and Long, E.O. (1985) DOB; a new *B* chain gene with a distinct regulation of expression. EMBO J. 4, 2839-2847

Tosi, G., Brunelli, S., Mantero, G., Magalini, A.R., Soffiati, M., Pinelli, L., Tridente G., Accolla, R.S. (1994) The complex interplay of the DQB1 and DQA1 loci in the generation of the susceptible and protective phenotype for insulin-dependent diabetes mellitus. Mol. Immunol. 31, 429-437

Townsend, A., Gotch, F.M, and Davey, J. (1985) Cytotoxic T cells recognise fragments of the influenza nucleoprotein. Cell 42, 457-467

Townsend, A., Rothbard, J., Gotch, F.m., Bahadur, G., Wraith, D. and McMichael A.J. (1986) The epitopes of influenza nucleoprotein recognised by cytotoxic T lymphocytes can be defined with short synthetic peptides. Cell 44, 959-968

Townsend, A., Ohlen, C., Bastin, J., Ljunggren, H-G., Foster, L. and Karre, K. (1989) Association of class-I major histocompatibility heavy and light chains induced by viral peptides. Nature 340, 443-448

Trowsdale, J. Young, J.A. Kelly, A.P. Austin, P.J. Carson, S. Meunier, H. So, A. Erlich, H.A. Spielman, R.S. Bodmer, J. et al. (1985) Structure, sequence and polymorphism in the HLA-D region. Immunol. Rev. 85, 5-43

Trowsdale, J., and Kelly, A.P. (1985) The human class ll *A* chain gene *DZA* is distinct from genes in the *DP, DQ* and *DR* sub-regions. EMBO J. 4, 2231-2237

Trowsdale, J., Groves, V., and Arnson, A. (1989) Limited MHC polymorphism in whales. Immunogenetics 29, 19-24

Trowsdale, J., Kelly, A., Lee, J., Carson, S., Austin, P and Travers, P. (1984) Linkage map of two HLA-SB beta and two HLA SB alpha-related genes, an intron in one of the SB genes contains a processed pseudogene. Cell 38, 241-249

Trowsdale, J., Lee, J., Carey, J. Grosveld, F., Bodmer, J and Bodmer, W. (1983) Sequence related to HLA-DR chain on human chromosome 6, restriction enzyme polymorphism detected with DC chain probes. Proc. Natl. Acad. Sci. USA 80, 1972-1976

Tsang S.Y., Nakanishi M., Peterlin B.M. (1990) Mutational analysis of the DRA promoter, cis-acting sequences and trans-acting factors. Mol. Cell. Biol. 10, 711-719

Unanue, E.R. (1984) Antigen presenting function of the macrophage. Annual Review of Immunology 2, 395-428

Van Der Poel, J.J., Groenen, M.A.M., Dijkhof, R.J.M., Ruyter, D. and Giphart, M.J. (1990) The nucleotide sequence of bovine MHC class II alpha genes DRA, DQA and DYA. Immunogenetics 35, 30-37

van Eijk, M.J. Stewart-Haynes, J.A. and Lewin, H.A. (1992c) Extensive polymorphism of the BoLA-DRB3 gene distinguished by PCR-RFLP. Animal Genetics 23, 483-496

van Eijk, M.J. Stewart-Haynes, J.A. Beever, J.E. Fernando, R.L. and Lewin, H.A. (1992b) Development of persistent lymphocytosis in cattle is closely associated with DRB2. Immunogenetics 37, 64-68

van Eijk, M.J.T., Russ, I. and Lewin, H.A. (1993) Order of bovine *DRB3, DYA* and PRL determined by sperm typing . Mammalian Genome 4, 113-118

van Eijk, M.J.T., Russ, I. Beever, J.E. and Lewin, H.A. (1992a) Polymorphism in exon 2 of the bovine lymphocyte antigen (BoLA) DYA gene. Animal Genetics 23, 476

Van Oorschot, R.A.H., Maddox, J.F., Adams, L.J. and Fabb, S.A. (1994) Characterisation and evolution of ovine MHC class ll *DQB* sequence polymorphism. Animal Genetics 25, 417-424

Van Rood, J.J., Eernisse, J.G., and Van Leeuwen, A. (1958) Leukocyte antibodies in sera from pregnant women. Nature 181, 1735

Vidovic D., Dembic Z. Qa-1 restricted gamma delta T cells can help B cells. (1991) Curr Top Microbiol Immunol. 173, 239-244

Viville S, Neefjes J, Lotteau V, Dierich A, Lemeur M, Ploegh H, Benoist C, Mathis D. (1993) Mice lacking the MHC class II-associated invariant chain. Cell 72, 635-648

Vogelstein, B. and Gillespie, D. (1979) Preparative and analytical purification of DNA from agarose. Proceedings of the National Academy of Sciences USA 76, 615-619.

Voliva, C.F., Tsang, S. and Peterlin, B.M., (1993) Mapping cis-acting defects in promoters of transcriptionally silent *DQA2, DQB2* and *DOB* genes. Proc Natl Acad Sci. USA 90, 3408-3412

Wake C.T., and Flavell R.A. (1985) Multiple mechanisms regulate the expression of murine immune response genes. Cell 42, 623-628

Watanabe, Y., Tsukada T., Notake M., Nakanishi S., Numa S., (1982) Structural analysis of repetitive DNA sequences in the bovine corticotropin-beta-lipotropin precursor gene region. Nucleic Acids Res. 10, 1459-1469

Wickens, M. (1990) How the messenger got its tail, addition of poly(A) in the nucleus. Trends in Biochem. Sci. 15, 277-281

Widera G. and Flavell R.A. (1984) The nucleotide sequence of the murine I-E beta b immune response gene, evidence for gene conversion events in class II genes of the major histocompatibility complex. EMBO J. 3, 1221-1225

Widera G., and Flavell R.A. (1985) The I region of the C57BL/10 mouse, characterization and physical linkage to H-2K of an SB beta-like class II pseudogene, psi Aβ3. Proc. Natl. Acad. Sci. USA 82, 5500-5504

Wigler, M., Pellicer, A., Silverstein, S., Axel, R., Urlaub, G. and Chasin, L. (1979) DNA-mediated transfer of the adenine phosphoribosyl transferase locus into mammalian cells. Proc. Natl. Acad. Sci. USA 76, 1373-1376

Williams A.F., and Barclay A.N. (1988) The immunoglobulin superfamily domains for cell surface recognition. Annual Review of Immunology 6, 381-405

Wiman, K., Larhammar, D., Claesson, L., Gustafsson, K., Schenning, L., Bill, P., Bohme, J., Denaro, M and Dobberstein, B. (1982) Isolation and identification of a cDNA clone corresponding to an HLA-DR antigen B chain. Proc. Natl. Acad. Sci. USA 79, 1703-1707

Wordsworth, B.P., Lanchbury, L.S.S., Sakkas, L.I., Welsh, K.I., Panayi, G.S and Bell, J.I. (1989) HLA-DR subtype frequencies in rheumatoid arthritis indicate that DRB1 is the major susceptibility locus within the HLA class-II region. Proc. Natl. Acad. Sci. USA 86, 10049-10053

Wraith, D.C., McDevitt, H.O., Steinman, L and Acha-Orbea, H. (1989) T-cell recognition as the target for immune intervention in autoimmune disease. Cell 57, 709-715

Wright H. and Ballingall K.T. (1994) Mapping and characterisation of the *DQ* sub-region of the ovine MHC. Animal Genetics 25, 243-249

Wright, H., Ballingall, K.T. and Redmond, J. (1994) The *DY* sub-region of the sheep MHC contains an A/B gene pair. Immunogenetics 40, 230-234

Wright, H., Redmond, J., Wright, F. and Ballingall, K.T. (1995) The nucleotide sequence of the sheep MHC class ll *DNA* gene. Immunogenetics 41, 131-133

Wright, H., Redmond, J. and Ballingall, K.T. (1996) The sheep orthologue of the *HLA-DOB* gene. Immunogenetics 43, 76-79

Xu A., Park C., Lewin H.A., (1994) Both DQB genes are expressed in BoLA haplotypes carrying a duplicated DQ region. Immunogenetics 39, 316-321

Xu A, Clark TJ, Teutsch MR, Schook LB and Lewin HA (1991) Sequencing and genetic analysis of a bovine DQB cDNA clone. Animal Genetics 22, 381-398

Xu A, McKenna K, Lewin HA (1993) Sequencing and genetic analysis of a bovine DQA cDNA clone. Immunogenetics 37, 231-234

Xu A, van Eijk MJ, Park C, Lewin HA (1993) Polymorphism in BoLA-DRB3 exon 2 correlates with resistance to persistent lymphocytosis caused by bovine leukemia virus. Journal of Immunology 151, 6977-6985

Xu Y., Pitcovski J., Peterson L., Auffray C., Bourlet Y., Gerndt B.M., Nordskog A.W., Lamont S.J., Warner C.M., (1989) Isolation and characterization of three class II MHC genomic clones from the chicken. Immunology 142, 2122-2132

Yanisch-Perron, C., Vieira, J and Messing, J. (1985) Improved M13 phage cloning vectors and host strains, nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33, 103-119.

Young, J.A.T. and Trowsdale, J. (199)) The HLA-DNA (DZA) gene is correctly expressed as a 1.1 kb mature mRNA transcript. Immunogenetics 31, 386-388

Zehetner G., Frishauf A and Lehrach H. (1987) Approach to restriction map determination. Nucleic Acid and Protein Sequence Analysis, A practical Approach (ed by M.J Bishop and C.J Rawlings) 147-164. ILR Press, Oxford.

Zinkernagel, R.M and Doherty, P.C. (1974) Restriction of in vitro T cell mediated cytotoxicity in lymphocytic choriomemingitis within a syngeneic or semiallogeneic system. Nature 248, 701-702

Zinkernagel, R.M. and Doherty, P.C. (1979) MHC-restricted cytotoxic T cells, studies on the biological role of polymorphic major transplantation antigens determining T cell restriction-specificity, function and responsiveness. Advances in Immunology 27, 51-177

Appendix 1. Phylogenetic trees of the MHC class ll *A* and *B* genes.

Phylogenetic trees constructed from the amino acid translations of the second and third exons from representative human, mouse, cattle, rabbit, and sheep MHC class ll *A* and *B* genes are shown in figures A1.1, 2.2, and 2.3. The trees were constructed by applying the neighbour-joining method (Saitou and Nei 1987) to pairwise distances that had been corrected for multiple hits by the method of Kimura (1983). Statistical testing of the tree topology was done using the bootstrap method based on 2000 trials (Hedges 1992). The bootstrap results are summarised as the percentage of trials supporting a given node. For example, the (*HLA-DNA* (*H-2OA, Ovar-DNA*)) cluster in figure A2.1 was supported by 99.8% of trials, suggesting that this cluster was indeed distinct from the rest of the tree. Only bootstrap percentages of over 95% are reported.

The PHYLIP computer program package was used for all phylogenetic analyses.

Legend to figure A1.1

Phylogenetic tree of MHC class ll *A* genes.

Sequences used in the tree construction were;

Sheep *Ovar-DQA1,* cosmids 4, 381 and 314,
Ovar-DQA2, Acc. No. m93433 and cosmids 2/9.2 and 22.1;
Ovar-DRA, Acc. No. z11600 (chapter 4).
Ovar-DYA, Acc. No. z27398 (chapter 6).
Ovar-DNA, Acc. No. z29533 (chapter 5).

Cattle *Bota-DYA*, Acc. No. m30119.
Bota-DRA, Acc. No. m30120),

Mouse *H-2Oa*, Acc. No. m95514;
I-E, Acc. No. j00396;
I-A, Acc. No. k01923).

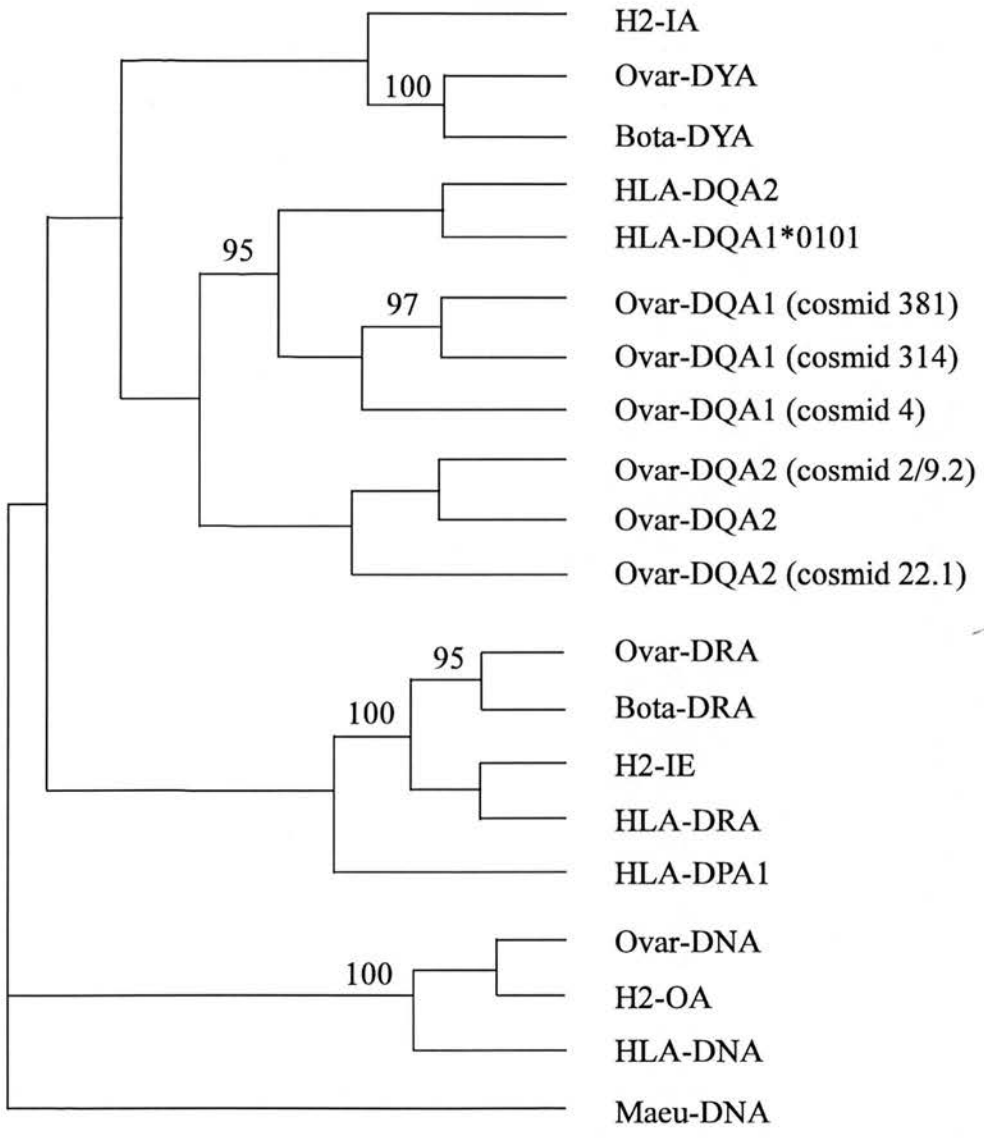
Human *HLA-DNA*, Acc. No. x02882;
HLA-DPA1, Acc. No. x03100;
HLA-DRA, Acc. No. j00204;
*HLA-DQA1*0101*, Acc. No. s40624
HLA-DQA2, Acc. No. m17237.

Tammar wallaby *Maeu-DNA*, Acc.No. l12121.

Figure A1.1

Phylogenetic trees of MHC class ll A genes.

See facing page for legend to this figure.



Legend to figure A1.2

Phylogenetic tree of translated second exons from sheep MHC class ll *B* genes together with representative genes from other species. Sequences used in the tree construction were as follows.

Sheep *Ovar-DQB1*, cosmids 314, 381, and 39.1 (chapter 3).
Ovar-DQB2, cosmids 309, 62 (chapter 3).
Ovar-DRB1, cosmid 22 and three alleles from Schwaiger et al. 1993
Ovar-DRB2, cosmid 9.1 (chapter 4).
Ovar-DRB4, cosmid 23.1 (chapter 4).
Ovar-DOB, cosmid 305 (chapter 5)
Ovar-DIB, cosmid 365 (chapter 6)

Human *HLA-DOB*, (Tonnelle et al. 1985).
*HLA-DQB1*0101*, (Marsh and Bodmer 1993)
HLA-DQB2, (Jonsson et al. 1987)
HLA-DPB1, (Marsh and Bodmer 1993)
*HLADRB1*0101* (Marsh and Bodmer 1993)

Cattle *Bota-DIB*, (Muggli-Cockett and Stone 1990)
*Bota-DQB2**3, (Xu et al. 1994)
*Bota-DQB1**2.0, (Xu et al. 1994)
Bota-DRB1, (Muggli-Cockett and Stone 1988)
Bota-DRB2, (Muggli-Cockett and Stone 1989)
Bota-DRB10, (Xu et al. 1994)

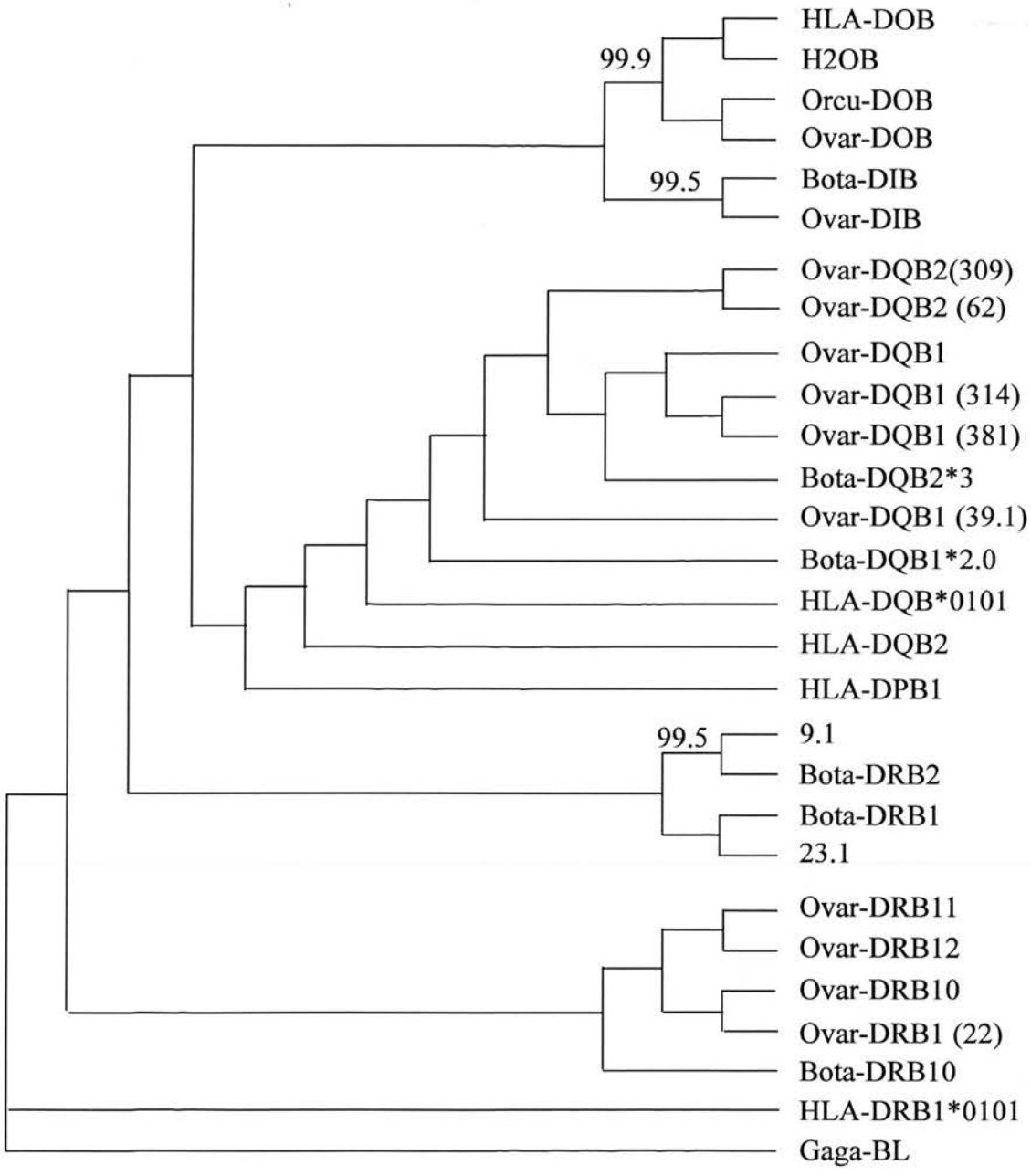
Mouse *H-2Ob*, (Larhammar et al. 1985)

Rabbit *Orcu-DOB*, (Chouchane et al. 1993).

Fowl *Gaga-BL*, (Xu et al. 1989)

Figure A1.2

Phylogenetic of MHC class ll *B* genes. See legend on facing page for details.



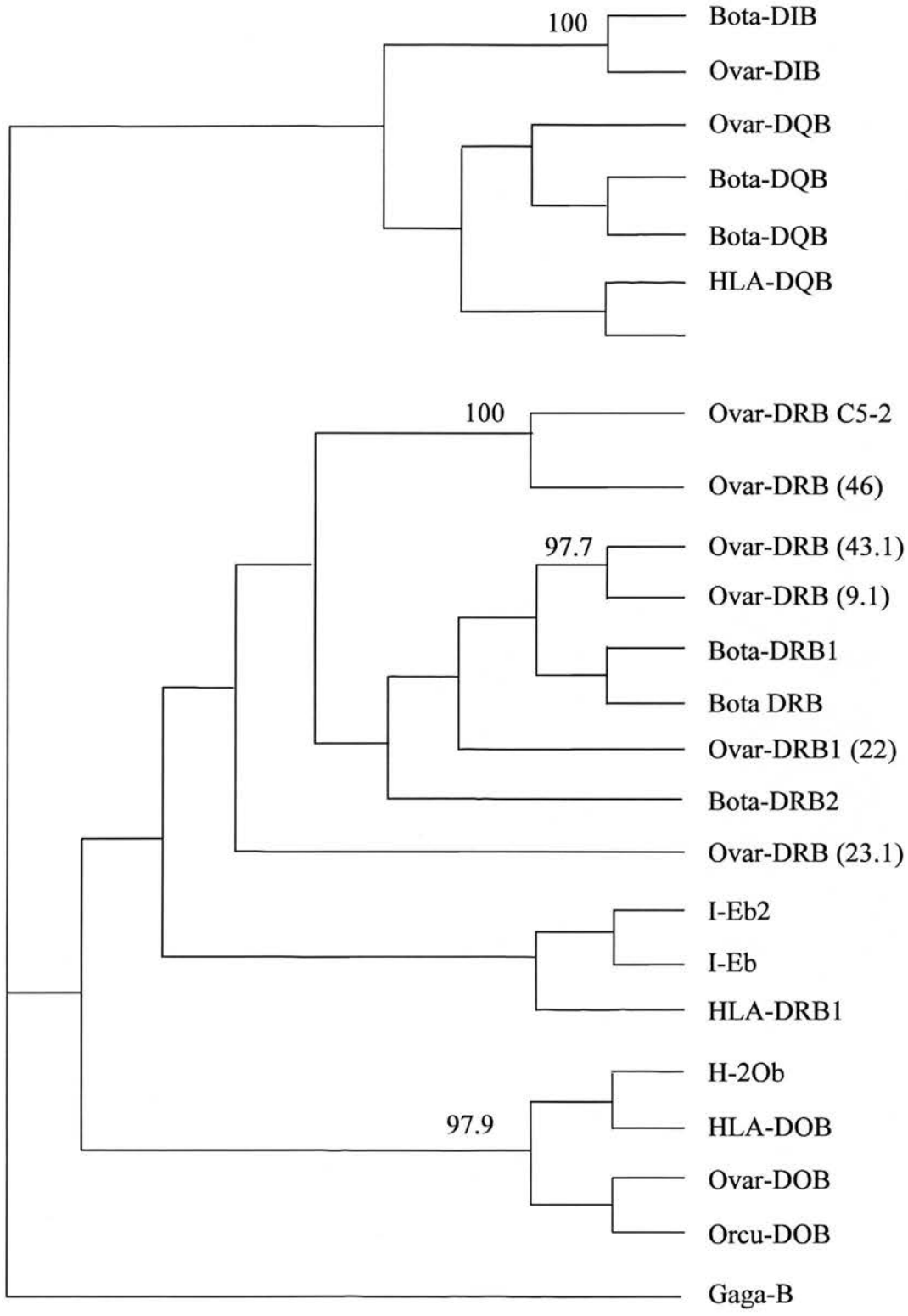
Legend to figure A1.3

Phylogenetic tree of translated third exons from sheep MHC class ll *B* genes together with representative genes from other species. Sequences used in the tree construction were as follows.

Sheep	*Ovar-DQB1*,	(Scott et al. 1990)
	Ovar-DQB2,	cosmid 62 (chapter 3).
	Ovar-DRB1,	cosmid 9.5/22
	Ovar-DRB2,	and *Ovar-DRB3* cosmid 9.1 (chapter 4).
	Ovar-DRB4,	cosmid 23.1 (chapter 4).
	Ovar-DRB5,	cosmid 46 (chapter 4).
	Ovar-DOB,	cosmid 305 (chapter 6)
	Ovar-DIB,	cosmid 365 (chapter 6)
Human	*HLA-DOB*,	(Tonnelle et al. 1985).
	*HLA-DQB1*0101*,	(Marsh and Bodmer 1993)
	*HLADRB1*0101*,	(Marsh and Bodmer 1993)
Cattle	*Bota-DIB*,	(Muggli-Cockett and Stone 1990)
	Bota-DQB,	(Groenen et al. 1990)
	Bota-DQB	
	Bota-DRB1,	(Muggli-Cockett and Stone 1988)
	Bota-DRB2,	(Muggli-Cockett and Stone 1989)
	Bota-DRB,	(Groenen et al. 1990)
Mouse	*H-2Ob*,	(Larhammar et al. 1985)
	I-Eb,	(Mengle-Gaw and McDevitt 1982)
	I-Eb2,	(Denaro et al. 1985)
Rabbit	*Orcu-DOB*,	(Chouchane et al. 1993).
Fowl	*Gaga-BL*,	(Xu et al. 1989)

Figure A1.3

Phylogenetic tree of the amino acid translations of the third exons from selected class ll *B* genes. See legend on facing page and text for details of tree construction.



Appendix 2. Publications arising from this thesis.

Ballingall, K.T., Wright H., Redmond J., Dutia B.M., Hopkins J., Lang J., Deverson E.V., Howard J.C., Puri N. and Haig D.M. (1992) Expression and characterization of ovine major histocompatibility complex class II (OLA-DR) genes. Animal Genetics 23, 347-359.

Wright H. and Ballingall K.T. (1994) Mapping and characterisation of the *DQ* sub-region of the ovine MHC. Animal Genetics 25, 243-249

Wright, H., Ballingall, K.T., and Redmond, J. (1994) The *DY* sub-region of the sheep MHC contains an A/B gene pair. Immunogenetics 40, 230-234

Wright, H., Redmond, J., Wright, F. and Ballingall, K.T. (1995) The nucleotide sequence of the sheep MHC class ll *DNA* gene. Immunogenetics 41, 131-133

Ballingall, K.T., Redmond, J. Dutia, B.M., Hopkins, J. and H. Wright. (1995) Analysis of the fine specificities of sheep major histocompatibility complex class ll-specific monoclonal antibodies using mouse L-cell transfectants. Animal Genetics 26, 79-84

Wright, H., Redmond, J., and Ballingall, K.T. (1995) The sheep orthologue of the *HLA-DOB* gene. Immunogenetics 43, 76-79

Copies are appended with permission of Springer-Verlag GmbH and Co. KG (Immunuogenetics) and of Blackwell Science Ltd. (Animal Genetics).

Immunogenetics (1994) 40: 230–234

BRIEF COMMUNICATION

Harry Wright · Keith T. Ballingall · James Redmond

The *DY* sub-region of the sheep MHC contains an *A/B* gene pair

Received: 23 November 1993 / Revised: 5 April 1994

Abstract The major histocompatibility complex (MHC) class II region of ruminants appears to have a structure broadly similar to that of the human class II or *HLA-D* region. Restriction fragment length polymorphism (RFLP) studies of class II genes in cattle (Andersson et al. 1988; Anderson and Rask 1988; Sigurdardottir et al. 1988, 1991 b), and in sheep (Scott et al. 1987), have provided an estimate of the number and type of class II genes in these species. The subsequent cloning and sequencing of sheep and cattle class II genes (Muggli-Cockett and Stone 1989; Groenen et al. 1990; van der Poel et al. 1990; Andersson et al. 1991; Scott et al. 1991 a, b; Ballingall et al. 1992; Sigurdardottir et al. 1991 a, 1992), have demonstrated that they are highly homologous to their human counterparts. Of more interest, therefore, are loci within the ruminant MHC which differ from the *HLA* class II region.

Three distinguishing features of the ruminant class II region described to date are, firstly, the apparent absence of a DP-like isotype, secondly, the variability in the number of DQ genes between haplotypes (Andersson and Rask 1988), and thirdly, the presence of class II genes presumed to be unique to the ruminant (Andersson et al. 1988). The presence of two such genes, designated *DYA* and *DYB*, was deduced from RFLP studies of cattle DNA. These genes were shown to segregate together with the *DOB* gene in one region separated by a recombination distance of 17 cM from the region which contains the *DQA, DQB, DRB, DRA*, and *C4* loci (Andersson et al. 1988). Subsequently, *Bota-DYA* was cloned from a phage library and sequenced (van der Poel et al. 1990; Acc. Nos. m30119 and m30118). The sequence of part of a similar gene in the goat, obtained by PCR by using primers derived from the cattle sequence, has recently been reported (Mann et al. 1993; Acc. No. m94325). However, there has been no report of the cloning of a *B* gene partner for the *DYA* gene. A novel cattle class II *B* gene designated *Bota-DIB* was cloned from a phage library and sequenced by Stone and Muggli-Cockett (1990). This was shown to be a single copy gene of limited polymorphism, which on the basis of RFLP analysis was probably not *Bota*-DYB but did appear to be distinct from other known cattle class II genes. The species distribution of this *B* gene was shown to be restricted to Cervidae, Giraffidae, and Bovidae (Stone and Muggli-Cockett 1993). However, it is not known whether any of these novel genes are functional.

Expressed human class II genes usually occur as *A/B* gene pairs situated close to each other on the chromosome. This is also the case with *Bota-DQ* genes (Groenen et al. 1990) and *Ovar-DQ* genes (Deverson et al. 1991; Wright and Ballingall 1994). We used the techniques of cosmid cloning and DNA-mediated gene transfection to determine whether there is a sheep equivalent of the *Bota-DYA* gene, whether there is a *DYB* gene partner, and whether there is a protein product.

A cosmid library was constructed from DNA prepared from a Finnish Landrace ram. The library was screened with *Ovar-DQA, Ovar-DQB, HLA-DQA*, and *HLA-DQB* gene probes at low stringency. A cosmid clone, 365, was obtained which hybridized weakly to both the *Ovar* gene probes. Restriction maps of the clone were produced for the enzymes *Eco* R1, *Bam* HI, *Hin* dIII, *Sac* I and *Sma* I. When the maps were compared to those published for the phage clones containing the *Bota-DYA* (van der Poel et al. 1990) and the *Bota-DIB* gene (Stone and Muggli-Cockett 1990), there was an imperfect match (Figure 1 shows the *Eco* RI maps). However, the sequence data for the *A* and *B* genes in cosmid 365 are more convincing. The sequences of exons 2 and 3 of the *A* gene in cosmid 365 and the *Bota-DYA* gene, together with the partial sequence from the third exon of the *Cahi-DYA* gene are shown in Figure 2 A. The predicted amino acid translations of these genes together with those of other published sheep MHC class II *A* genes are shown in Figure 2 B. The *A* gene in cosmid 365 had all the salient features of

The nucleotide sequence data reported in this paper have been submitted to the EMBL nucleotide sequence database and have been assigned the accession numbers z27398–z27399 (*DYA*) and z27400–z27401 (*DYB*)

H. Wright (✉) · K. T. Ballingall · J. Redmond
Moredun Research Institute, 408 Gilmerton Road, Edinburgh, EH17 7JH UK

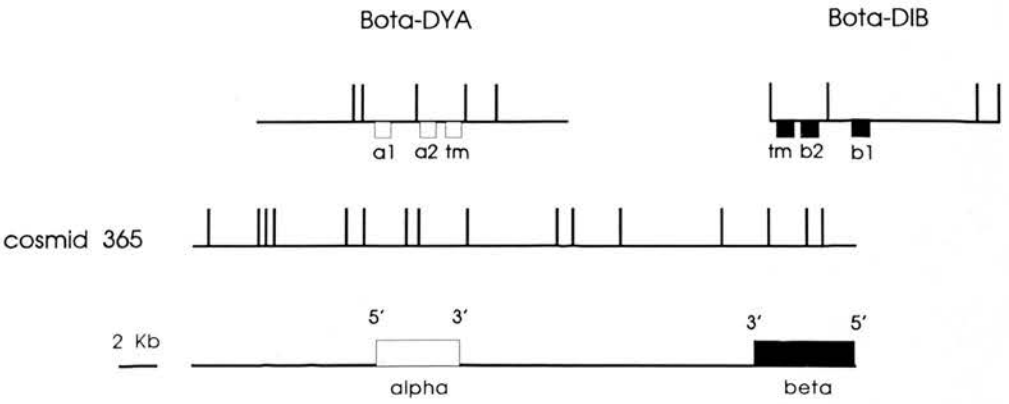


Fig. 1 *Eco* RI restriction maps of the sheep cosmid 365 and the phage clones of the cattle MHC class II *DYA* and *DIB* genes

A

```
                exon 2
Bota-DYA        agctGACCACGTGGGCACTTACGGCACAAATGTCTACCAGACGTACGGCACCTCTGGCCAGTTCACGTTTGAATTTGATGGAGACGAGCTCTTCTACGTG
Ovar-DYA        ---g----------------------------------------------------------G-----------------------------------

Bota-DYA        GACCTGGGGAAAAAGGAGACTGTCTGGCGGCTGTCCGAGTTTAGCAATATCACCAAGTTTGAAGTTCAGAGCGCCCTGAGAAACATTGTTATGTCAAAAA
Ovar-DYA        ------A--------------------------C---------A---------A-T-------A-------T----------------------------

Bota-DYA        GAAATTTGGACATCTTGATAAAAAATTCCAGCTTTACACCTGCCACCAGTGgt....1.19 kb....gaaataatacactaacagctggggagattgtg
Ovar-DYA        -------------------G----------A-----------------AC---                 ---------------g------------g---
                                                                                                        exon 3
Bota-DYA        cacggggtggggaagcctcccctaaactgattaaggagcagaaggcacagttgtcagaaattccaatccatctttggtcttcattgtagAAATCCCTGAA
Ovar-DYA        --t---------g-----------------g---------t-------g-----------------------------------------C---------

Bota-DYA        GTGGCTGTGTTTCCCAAATCCTCTGTGGTCCTGGGGATTCCCAATACCCTCATCTGTCAAGTGGACAACATCTTTCCTCCTGTGATCAACATCACTTGGT
Ovar-DYA        ------------------------C---------------------C-----------------------------------------------------
Cahi-DYA                                                        ---------------------------------------------------

Bota-DYA        TTTACAATGGACACTTTGTTGCAGAAGGGATCGCTGAGACCACCTTCTACCCCAAGAGTGACCACTCTTTCCTCAAGTTCAGTTACCTCACCTTTCTTCC
Ovar-DYA        -------------G--------------TG--A-----------------------------------C---------------------------G----
Cahi-DYA        -------------A--------------TG-------------------------------------C-----------------------------------

Bota-DYA        CACCAGTGAAGACTTCTATGACTGCAGAGTGGAGCACTGGGGCCTGGAAGAGCCCCTCGTCAAGCACTGGGgtacgtgcattcccaaacccacaccctc
Ovar-DYA        -G-----------------------------------------------G-------------------------------a----a---g------

Bota-DYA        tccacatccaatccacctgcagacagtgtccttccgaatccggcctcc
Ovar-DYA        ----------------------t------------t------a-----t
```

B

```
                exon 2
Bota-DYA        DHVGTYGTNVYQTYGTSGQFTFEFDGDELFYVDLGKKETVWRLSEFSNITKFEVQSALRNIVMSKRNLDILIKNSSFTPATS
Ovar-DYA        ---------------A-----------------R--------P--N---M--I----------------M---N-----N
Ovar-DQA1       --I----V-I-----P--YY-H------E ----E-R------P---KF-S-DP-G----IATV-H--E---QR-NS-A--N
Ovar-DQA2       --F-S---EI--S -P---Y-Q----------------------PM--QFAG-DP-G--SE-ATA-Q-----T-R-N----IN
Ovar-DRA        ---*IIQAEF-LNPEE-AE-M-D-----I-H--MQ--------P--GHFAS--A-G--A-MAVM-A----M--R-NN--N-N

                exon 3
Bota-DYA        EIPEVAVFPKSSVVLGIPNTLICQVDNIFPPVINITWFYNGHFVAEGIAETTFYPKSDHSFLKFSYLTFLPTSEDFYDCRVEHWGLEEPLVKHW
Ovar-DYA        D-----------------------------------------Q-----VT--------------------V-A---------------G------
Cahi-DYA                                 ---------------------Q-----V-----------------------
Ovar-DQA1        V---T--S--P-M--Q------H-------------LR---S-T--VS--S-LI---Y----IN------SDD-V---K------D---L---
Ovar-DQA2        V---T--S--P-M--Q------H-------------LK---A-T--VS--S-L--D------IG------SDD-V---K------D---L---
Ovar-DRA        **---TLL-NKP-E--E------FI-KFS----SV--LR--IP-TD-VSQ-V-L-RD--L-R--H--P----T--V---K------N---L---
```

Table 1 The % identity between exons 2 and 3 of the *Ovar-DYA* gene and those of the *Bota-DYA, Cahi-DYA, Ovar-DQA1, Ovar-DQA2*, and *Ovar-DRA* genes. Asterisks indicate that only a partial sequence is available for the caprine gene. Figures in parenthesis are the % similarity between the amino acid sequences

	Nucleotides		Protein	
	Exon 2	Exon 3	Exon 2	Exon 3
Bota-DYA	95	96	89 (92)	92 (98)
Cahi-DYA	*	98*	*	96*(98)
Ovar-DQA1	75	76	67 (82)	69 (83)
Ovar-DQA2	74	78	66 (76)	71 (85)
Ovar-DRA	68	68	54 (71)	58 (72)

Fig. 2 **A** Comparison of the nucleotide sequences of exons 2 and 3 of the *Bota-DYA* gene (van der Poel et al. 1990) and the *A* gene in cosmid 365. The partial sequence of exon 3 of the goat gene (Mann et al. 1993) is also shown. Exon sequences are in *uppercase*. The splice junctions are *underlined. Dashes* indicate identity. **B** Comparison of the amino acid sequences derived from the nucleotide sequences for exons 2 and 3 of the *Bota-DYA* gene, the *A* gene in cosmid 365, *Ovar-DQA1* and *DQA2* genes (Scott et al. 1991 a), and the *Ovar-DRA* gene (Ballingall et al. 1992). The partial sequence of the goat exon 3 is also shown. *Asterisks* indicate residues not present in the DR isotype

A

```
                exon 2
Bota-DIB    AGAGAATTTCGTGTACCAGTTTAAAGGCATGTGCTACTTCACCAACGGGACAGAGCACGTGAGGCTTGTGGCCAGACAGATCTACAACAAGGAAGAGATCCTGCACTTTG
Ovar-DYB                                                                                                        G----------

Bota-DIB    ACAGTGACCTGGGCGAGTTTGTGGCTGTTACAGAGCTGGGCCGGGTGTGTGCGGAGATCTGGAACACCCAGAAGGACCTCCTGGCGGAGTTTCGGGCCTACGTGGACACG
Ovar-DYB    ----C-------------------------------------------------C--T-------------------------------------------------G---
                                                                                                          exon 3
Bota-DIB    CTGTGTAGACACAACTACAAAGAGACGGCCGGCTTCACTGTCCAGCGGAGAGgt....2.7kb....cctcattctctaattttgtctcgtttcctcctagTGGAGCCTA
Ovar-DYB    --------------------G-----------------------A--------            -g----------g---------t--------------------

Bota-DIB    CAGTGACTGTCTCTCCAGCCAGTACAGAGGCCCTGAACCACCATAATCTGCTGGTCTGTTCAGTGACAGATTTCTACCCTCGCCAAGTTAAAGTCAAATGGTTCCGGAAT
Ovar-DYB    -------G-------------------------C-----T-----------------------------A-------T---------------A----------------

Bota-DIB    CAACAGGAGCAGACAGCTGGAGTTGGGTTCACACCTCTTACTCAGAATGGGGACTGGACCTACCAGATTCACGTGATGCTAGAGACAGTTCCACAGCTTGGAGACGTCTA
Ovar-DYB    ------------------------------------------------------------------------------------A-A-------------------

Bota-DIB    CGTTTGCCACCTGGACCACCCCAGCCTCCAGAGCCCCATCACAGTAGAATGGCgtaagggccgcttcactgaccctacggacccgacaggaaagaaaaagttcagggaga
Ovar-DYB    T---------G-------------------------------------------------------t---.---------t------a---------a----------

Bota-DIB    gcgctgggtctggtgggggtggactccgtcttcatccctcgtgtcctatgtaactccctgatacaatttctgggctggaagtgaccgaggactagatcccagtat.....
Ovar-DYB    -t---------------------g----------------a------g------c-------------------------------------------------ctcaa

Bota-DIB    .........catctgttgagtccttatctcatttcccccccccaagatgtgatggtggtcccttcacacgaccggaccccccccccc   50bp not sequenced in
Ovar-DYB    gttgaaagg------c-------------------..t------------g-------------------t--t--------t-t---

Bota-DIB    the bovine   gggctttgacgtagaggctggt.atctgaa.gtgacccgtggaggacagacaaacagatggacatgccctcttgg..ggctaccagcct.cccccac
Ovar-DYB                 ----------a-----------c--------g-----------------------....-----------tc-----ga-a--g---t---c-----gt

Bota-DIB    catccatc.gagcccccccgg.gccttgcctgccctcccctcttctgtctggtttctgatcccctctgtctcttatacacacaaccagactccaggcttccagacgaggat
Ovar-DYB    ----t----t------t---at-tt----c--t-------------------------g----------------------tg----------------------a-----
                                                                                  exon 4
Bota-DIB    gctg..tgggccgtggggacactgacctgg..ggcttttacttccagGGGCACAGTCTGAATCAGCCCAGAGCAAGATGCAGAGCGGAATTGGAGGTTTTGTGCTGGGGC
Ovar-DYB    ----ata------------ac----gc---ga-------------------------------T-T--------------TG-----------------------------

Bota-DIB    TGATCTTCCTTGGTGTGGGCCTTTTTGTCCACTTTTGGGATAAGAGAGgtaaggcgccttgggagaaaaaggggggaagacaggtgctgggctgaaaagcctctgttgatc
Ovar-DYB    ------------------------------------CA--------------g---a-------------.--------------c---a----------------------

Bota-DIB    cttctctagtgactgtctcagtgacgttggggtgatcttcttttgtggcagttgaatccttattgtatgtgggtgggagcaaagaacgctctggaactgccccccctcatt
Ovar-DYB    ------------------------a--------g-------------t-------------------------------------------t----------------------
```

B

```
                exon 2
Bota-DIB    NFVYQFKGMCYFTNGTEHVRLVARQIYNKEEILHFDSDLGEFVAVTELGRVCAEIWNTQKDLLAEFRAYVDTLCRHNYKETAGFTVQRR
Ovar-DYB                                                                                    E-----------------
Ovar-DQB    D--FL-M-Q--------R----T-Y---Q--H-R----W--YR---PP-QRQ--YF-S---I-ERTG-EA--V-----QVE-A--W---
Ovar-DRB    H-LEYT-KE-R-S----R--FLD-YF--G--YVR----W--YR--A----PD-KY--S--EI-ERR-TE---Y-----GVIES-S----

                exon 3
Bota-DIB    VEPTVTVSPASTEALNHHNLLVCSVTDFYPRQVKVKWFRNQQEQTAGVGFTPLTQNGDWTYQIHVMLETVPQLGDVYVCHLDHPSLQSPITVEW
Ovar-DYB    --------------P-L---------N-------I-------------------------------------I----------V-------------
Ovar-DQB    ---------SR--------------------G-I--R----DR-E----VS---IR-----F--L----MT--R----T-RVE--------S---
Ovar-DRB    ---I---Y--K-QP-Q---------NG---GHIE-R----GH-EE---IS-G-I------F-TM--------S-E--T-QV----RT-------

                exon 4
Bota-DIB    RAQSESAQSKMQSGIGGFVLGLIFLGVGLFVHFWDKR
Ovar-DYB    ------V----W----------------------Q---
Ovar-DQB    -----------L--V----------SL--II-HRSQK
Ovar-DRB    --R-D------M--V-------L--A----IY-R-QK
```

Fig. 3 **A** Comparison of the nucleotide sequences of exons 2, 3, and 4 of the *Bota-DIB* gene (Stone and Muggli-Cockett 1990) and the *B* gene in cosmid 365. The sheep sequence starts midway through exon 2. Gaps have been introduced into intronic sequences to maintain alignment and are indicated by *dots*. *Dashes* indicate identity. Exon sequences are in *uppercase*. Splice junctions are *underlined*. **B** Comparison of the amino acid sequences derived from the nucleotide sequences of exons 2, 3, and 4 of the *Bota-DIB* gene, the *B* gene in cosmid 365, an *Ovar-DQB* gene (Scott et al. 1990), and an expressed *Ovar-DRB* gene (Ballingall et al. 1992). *Dashes* indicate identity

an MHC class II A gene. It showed a high sequence similarity to the cattle and caprine *DYA* genes and much less so to the *Ovar-DRA* gene (Ballingall et al. 1992; Acc. No z11600) and the *Ovar-DQA1* and *DQA2* (Scott et al. 1991a; Acc. Nos. m33304 and m33305), as detailed in Table 1. The cosmid *A* gene showed low sequence similarity to the sheep *DNA* (formerly *DZA*) gene (unpublished observations). The *A* gene described here is clearly the sheep homologue of the *Bota-DYA* gene.

The sequences of the second, third, and fourth exons of the B gene in cosmid 365 are shown in Figure 3A together with those of the *Bota-DIB* gene (Stone and Muggli-Cockett 1990). Unfortunately, the presence of a *Bam* HI site in exon 2 of the sheep gene caused a truncation at this point, during the cloning procedure and so a part of exon 2, the whole of exon 1, and all the upstream regulatory elements were missing. The predicted amino acid translations of exons 2, 3, and 4 are shown together with those of an *Ovar-DQB* (Scott et al. 1991a; Acc. No. m33323) and an expressed *Ovar-DRB* gene (Ballingall et al. 1992; Acc. No. z11522) in Figure 3B.

Table 2 The % identity between exons 2, 3, and 4 of the *Ovar-DYB* gene and those of *Bota-DIB, Ovar-DQB*, and *Ovar-DRB* at both the nucleotide and amino acid levels. Figures in parenthesis are the % similarity between the amino acid sequences

	Nucleotides			Protein		
	Exon 2	Exon 3	Exon 4	Exon 2	Exon 3	Exon 4
Bota-DIB	96	96	95	98 (100)	94 (98)	92 (92)
Ovar-DQB	71	78	81	49 (64)	73 (83)	66 (86)
Ovar-DRB	71	72	74	53 (68)	62 (75)	68 (84)

The percent identities between the sheep gene and the *Bota-DIB, Ovar-DQB*, and *Ovar-DRB* genes for each exon at both the nucleotide and amino acid levels are shown in Table 2. The *B* gene in cosmid 365 has all the characteristics of an MHC class II *B* gene. It showed a high degree of sequence similarity to *Bota-DIB* and was quite different from the *Ovar-DQB* and *DRB* genes, particularly in the amino acid translations of the second exon. We feel justified in calling the *B* gene in cosmid 365 *Ovar-DYB* because of its proximity to the *A* gene which is clearly *Ovar-DYA*. The distance between the genes is about 11 kilobases (kb) and the transcriptional orientation is tail-to-tail, a structure similar to that of the *HLA-DQ* sub-region (Campbell and Trowsdale 1993), the *Bota-DQ* sub-region (Groenen et al. 1990), and the *Ovar-DQ1* genes (Deverson et al. 1991; Wright and Ballingall 1994).

We have as yet no estimate of the distance between the *DY* locus and the other class II loci in the sheep. Although there is no estimate of the distance between *Bota-DYA* and *Bota-DRB3* in terms of kb of DNA, the genetic distance in terms of recombination frequency has been estimated at 15 cM (van Eijk et al. 1993), a figure that agrees well with the 17 cM estimated for the distance between the two sub-regions containing *DY/DO* on the one hand and *DQ/DR* on the other (Andersson et al. 1988). These frequencies contrast with the estimated recombination distance of only 3 cM between the *HLA-DQ* and *HLA-DP* sub-regions which are separated by 300 kb on the molecular map (Campbell and Trowsdale 1993). Stone and Muggli-Cockett (1993) have shown that the *Bota-DIB* gene maps to the group containing the *DOB* gene and hence, by inference, the *DYA* gene. This is in accord with the evidence presented here.

The Ovar-DR and Ovar-DQ isotypes have been expressed at the surface of mouse L cells following transfection with cosmid DNA (Ballingall et al. 1992; Wright and Ballingall 1994). The same technique was applied, using the *DY* cosmid 365 without success. This was perhaps not surprising when it was realized that the *B* gene was truncated (Fig. 1). However, evidence for transcription of the *DYA* gene was obtained when total RNA prepared from the transfected L cells was hybridized with a *DYA* probe (data not shown). Stone and Muggli-Cockett (1993) were unable to detect significant amounts of *Bota-DIB* mRNA in Northern blots of peripheral blood monocytes. Whether an Ovar-DY class II molecule is expressed at the cell surface remains to be demonstrated. A phylogenetic analysis of *Bota A* genes by Scott and co-workers (1991a) placed the *DYA* gene on the same branch as the *DQA* genes but indicated that *DYA* branched off long before the duplication which led to *DQA1* and *DQA2*. The numerous mutations which have accumulated in the *Bota-DYA* gene since it branched off from the *DQ* line led the same authors to conclude that the *DY* genes were unlikely to be functional. On the other hand, sheep and cattle shared a common ancestor 15–20 million years ago and the high degree of sequence similarity between the sheep *DY* genes and the cattle *DYA* and *DIB* genes provides an argument that these are evolutionary conserved MHC genes of unknown function.

In summary, we have shown that the *Ovar-DY* locus is represented by an *A/B* gene pair, a structure similar to that of the *Ovar-DQ1* locus. In terms of the three possibilities cited by Andersson and co-workers (1988) to account for the *DY* locus, namely, duplication and transposition of *DQ* or *DR* genes, gene conversion of *DP* genes, or lastly, new class II genes specific to ruminants, the data presented here would favor the duplication of a pair of *DQ* genes, with subsequent rapid divergence.

Acknowledgments This work benefited from the use of the SEQNET facility. The authors are supported by funds from the Scottish Office Agriculture and Fisheries Dept. The clones are freely available to interested parties.

References

Andersson, L., Lunden, A., Sigurdardottir, S., Davies, C. J., and Rask, L. Linkage relationships in the bovine MHC region: High recombination frequency between class II subregions. *Immunogenetics 27:* 273–280, 1988

Andersson, L. and Rask, L. Characterisation of the MHC class II region in cattle. The number of *DQ* genes varies between haplotypes. *Immunogenetics 27:* 110–120, 1988

Andersson, L., Sigurdardottir, S., Borsch, C., and Gustafsson, K. Evolution of MHC polymorphism: extensive sharing of polymorphic sequence motifs between human and bovine *DRB* alleles. *Immunogenetics 33:* 188–193, 1991

Ballingall, K. T., Wright, H., Redmond, J., Dutia, B. M., Hopkins, J., Lang, J., Deverson, E. V., Howard, J. C., Puri, N., and Haig, D. M. Expression and characterisation of ovine major histocompatibility complex class II (OLA-DR) genes. *Anim Genet 23:* 347–359, 1992

Campbell, R. D. and Trowsdale, J. A map of the human MHC. *Immunol Today 14:* 347–352, 1993

Deverson, E. V., Wright, H., Watson, S., Ballingall, K. T., Huskisson, N., Diamond, A. G., and Howard, J. C. Class II major histocompatibility complex genes of the sheep. *Anim Genet 22:* 211–225, 1991

Groenen, M. A. M., Van der Poel, J. J., Dijkhof, R. J. M., and Giphart, M. J. The nucleotide sequence of the bovine MHC class II *DQB* and *DRB* genes. *Immunogenetics 31:* 37–44, 1990

Mann, A. J., Abraham, L. J., Cameron, P. U., Robinson, W., Giphart, M. J., and Dawkins, R. L. The caprine MHC contains *DYA* genes. *Immunogenetics 37:* 292–295, 1993

Muggli-Cockett, N. E. and Stone, R. T. Partial nucleotide sequence of a bovine major histocompatibility class II *DRB*-like gene. *Anim Genet 20:* 361–369, 1989

Scott, P. C., Choi, C.-L., and Brandon, M. R. Genetic organisation of the ovine MHC class II region. *Immunogenetics 25:* 116–122, 1987

Scott, P. C., Gogolin-Ewens, K. J., Adams, T. E., and Brandon, M. R. Nucleotide sequence, polymorphism and evolution of ovine MHC class II *DQA* genes. *Immunogenetics 34:* 69–79, 1991a

Scott, P. C., Maddox, J. F., Gogolin-Ewens, K. J., and Brandon, M. R. The nucleotide sequence and evolution of ovine MHC class II *B* genes: *DQB* and *DRB. Immunogenetics 34:* 80–87, 1991 b

Sigurdardottir, S., Lunden, A., and Andersson, L. Restriction fragment length polymorphism of *DQ* and *DR* class II genes of the bovine major histocompatibility complex. *Anim Genet 19:* 133–150, 1988

Sigurdardottir, S., Borsch, C., Gustafsson, K., and Andersson, L. Cloning and sequencing of 14 *DRB3* alleles of the bovine MHC by using the polymerase chain reaction. *Anim Genet 22:* 199–210, 1991 a

Sigurdardottir, S., Mariani, P., Groenen, M. A. M., Van der Poel, J. J., and Andersson, L. Organisation and polymorphism of bovine major histocompatibility complex class II genes as revealed by genomic hybridisations with bovine probes. *Anim Genet 22:* 465–476, 1991 b

Sigurdardottir, S., Borsch, C., Gustafsson, K., and Andersson, L. Gene duplications and sequence polymorphism of bovine class II *DQB* genes. *Immunogenetics 35:* 205–213, 1992

Stone, R. T. and Muggli-Cockett, N. E. Partial nucleotide sequence of a novel bovine major histocompatibility complex class II beta chain gene, *BoLA DIB. Anim Genet 21:* 352–360, 1990

Stone, R. T. and Muggli-Cockett, N. E. *BoLA-DIB:* species distribution, linkage with *DOB* and Northern analysis. *Anim Genet 24:* 41–45, 1993

Van der Poel, J. J., Groenen, M. A. M., Dijkhof, R. J. M., Ruyter, D., and Giphart, M. J. The nucleotide sequence of the bovine MHC class II alpha genes: *DRA, DQA* and *DYA. Immunogenetics 31:* 29–36, 1990

Van Eijk, M. J. T., Russ, I., and Lewin, H. A. Order of bovine *DRB3, DYA* and PRL determined by sperm typing. *Mammalian Genome 4:* 113–118, 1993

Wright, H. and Ballingall, K. T. Mapping and characterisation of the DQ subregion of the ovine MHC. *Anim Genet,* in press

Animal Genetics, 1994, **25**, 243–249

Mapping and characterization of the *DQ* subregion of the ovine MHC

H Wright, K T Ballingall

Summary

A map of the ovine MHC class II *DQ* subregion has been constructed from overlapping cosmid clones. This region consists of two loci linked on a linear tract of 130 kb DNA. Each locus consists of a *DQA* and a *DQB* gene in a tail-to-tail orientation. The genes in each locus are transcribed but only those designated *DQ1* express class II molecules at the surface of mouse L cells following DNA-mediated gene transfection. The *DQA1* and *DQB1* genes are separated by 11 kb while the *DQA2* and *B2* genes are 25 kb apart. The loci are separated by 22 kb.

Keywords: ovine, MHC-DQ, subregion mapping, gene expression

Introduction

The products of the class II genes within the major histocompatibility complex are heterodimeric cell surface glycoproteins that present fragments of predominantly exogenous foreign proteins to CD4+ T lymphocytes (Germain & Margulies 1993). Extensive nucleotide polymorphisms within these genes are implicated in variations in immune responses within a species (Racioppi *et al.* 1991). Previous studies of the class II region of the sheep MHC (*OvarMHC*, Klein *et al.* 1990), have identified genes equivalent to all HLA subtypes with the exception of the HLA *DP* region (Chardon *et al.* 1985; Scott *et al.* 1987, 1991a, 1991b; Deverson *et al.* 1991; Ballingall *et al.* 1992). What is still not clear is how many class II genes there are; how they are arranged on chromosome 20 (Hediger *et al.* 1991); how many encode class II molecules or are pseudogenes or gene fragments, and what variation there is between haplotypes. In an attempt to answer at least some of these questions a contiguous map of the *Ovar-MHC* class II region from overlapping cosmid clones is being produced. The production and screening of two ovine genomic libraries using human and murine class II probes representing the major class II subtypes has been described previously by Deverson *et al.* (1991). A large number of *A* and *B* genes were identified, isolated and mapped. The *Ovar-DRA* and *B* genes were expressed following transfection of mouse L cell fibroblasts (Ballingall *et al.* 1992). However, despite the cloning of a number of *Ovar-DQA* and *B* genes we were unable to express any of these using this technique. Therefore, in an attempt to complete the map of the *DQ* subregion and identify expressible class II genes, a new cosmid library was produced. This paper describes the cloning, mapping, partial nucleotide sequence analysis, linking of *DQA1/B1* and *DQA2/B2* loci using overlapping cosmid clones and an evaluation of the ability of each of these genes to express cell surface proteins in mouse L cells.

H Wright
K T Ballingall
Moredun Research Institute, 408 Gilmerton Road, Edinburgh EH17 7JH, UK

Correspondence: Keith T Ballingall.

Accepted 21 April 1994

Materials and methods

The construction and screening of cosmid libraries

A cosmid library was prepared using the vector pCos8 (a derivative of the pCos6 vector described by Ehrlich *et al.* 1987). Vector arms were prepared and dephosphorylated according to standard procedures (Steinmetz *et al.* 1982). Sheep whole blood DNA was prepared from a male Finnish Landrace sheep using the procedure of Herrmann & Frischauf (1987). The DNA was partially restricted using the enzyme *Nde*II and size-fractionated by centrifugation on a NaCl gradient. Colonies ($7{\cdot}5 \times 10^5$) were plated out on three 20 × 20 cm plates and then washed off to make a cell suspension which was stored at −70°C. Colonies (6×10^5) were screened on three 20 × 20 cm Hybond-N membranes (Amersham International, Aylesbury, UK) using DQ-specific probes. The construction and screening of two additional libraries from which the *DQ* cosmids 2, 4, 9.2, 22.1, 39.1 and 62 were isolated and mapped have been described previously (Deverson *et al.* 1991).

Restriction maps

Cosmids were linearized with lambda terminase (Amersham International, Aylesbury, UK) and

mapped for *Bam*HI, *Hin*dIII, *Eco*RI, *Sac*I and *Sma*I as described by Zehetner *et al.* (1987).

Nucleotide sequencing

The second exons of the class II *DQ A* and *B* genes were identified using exon-specific probes. Following subcloning into M13mp18/19 the nucleotide sequence was determined by the dideoxy chain termination method of Sanger *et al.* (1977) using Sequenase 2.0 or Taquence enzymes together with the 7-deazadGTP analogue (United States Biochemicals, Ohio, USA) to overcome problems with compressions.

DNA-mediated gene transfer

Stable transfection of ovine class II genes into mouse L cells was as previously described (Ballingall *et al.* 1992). L cells expressing ovine class II molecules were identified by an indirect fluorescent antibody test using a cocktail of MHC class II-specific monoclonal antibodies and FACScan analysis. L cell lines were generated by cycles of sorting (FACStar, Becton Dickinson, Sunnyvale, CA, USA) until >90% of the cells expressed *Ovar-DQ* molecules. Transient expression of *Ovar-DQ* genes was achieved by transfection of mouse L cells with 20 µg cosmid DNA using a modification of the $CaPO_4$ method (Lang *et al.* 1983). Cells were glycerol shocked (15% glycerol in DMEM for 1 min) following 4 h contact with the DNA/$CaPO_4$ precipitate. After three washes in PBS, cells were cultured in DMEM and harvested for analysis 48 h later.

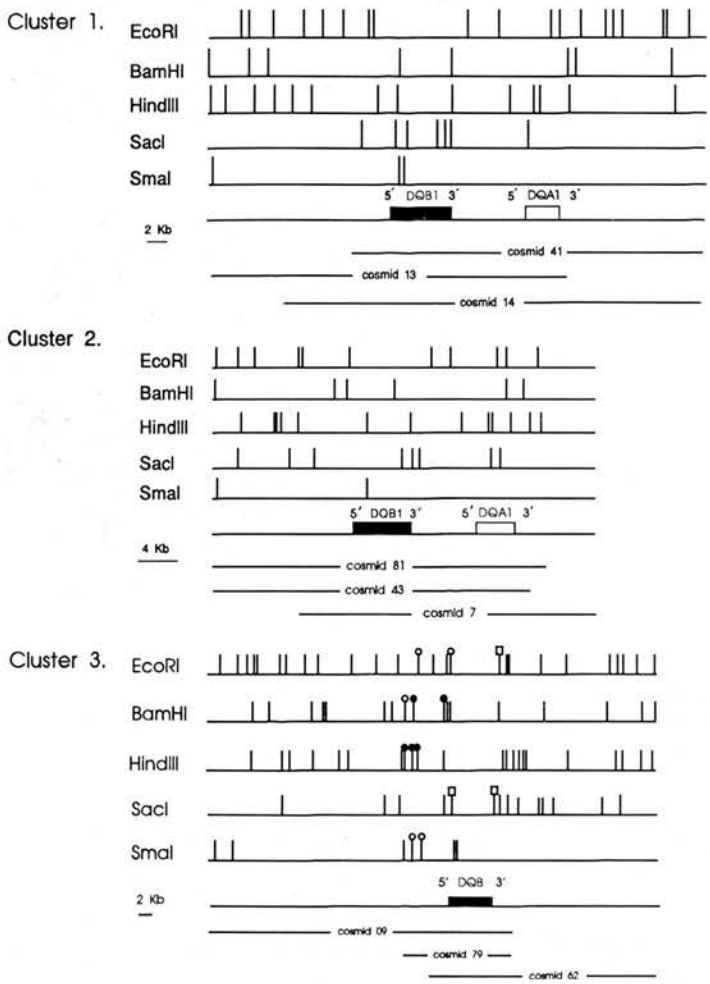


Fig. 1. Restriction maps of overlapping cosmids containing ovine MHC class II *DQ* genes. Clusters 1 and 2 represent two alleles of an *A/B* gene pair at the expressed *DQ1* locus. Cluster 3 represents three alleles at the *DQB2* locus. Open circles indicate sites unique to cosmid 79, closed circles indicate sites unique to cosmid 09 and squares indicate sites unique to cosmid 62.

Results

Restriction maps and nucleotide sequences

Six cosmid clones (7, 13, 14, 41, 43 and 81) were identified that each contained both an MHC class II *DQA* and *DQB* gene. Two additional cosmids, 9 and 79, each contained a single *DQB* gene. The group of six cosmids could be arranged into two clusters of three overlapping clones based on restriction mapping (Fig. 1, clusters 1 and 2). Within a cluster, restriction sites matched perfectly, whereas between clusters approximately 90% of the sites matched. Furthermore, within a cluster the nucleotide sequences of the second exons of both the *A* and the *B* genes were identical, while between clusters the sequences varied (Figs 2 and 3). The four differences in the nucleotide sequence between the *A* genes in the two clusters each resulted in a change in the predicted amino acid sequence. The sequence of the *A* gene in cluster 1 was identical to that published by Scott *et al.* (1991a) (EMBL Accession No. M33304) which they designated *DQA1*, while that of cluster 2 was 98% similar (Table 1). The *A* genes in clusters 1 and 2 were 78% similar to the *DQA2* sequence described by Scott *et al.* (1991a) (Accession No. M33305, Fig. 2 and Table 1). The sequences of the *DQB* genes in clusters 1 and 2 and those within the EMBL database were between 90 and 94% similar to one another (Fig. 3). The 21 nucleotide differences between the *B* genes in clusters 1 and 2 would result in 13 amino acid changes with only three silent substitutions.

The restriction maps of the cosmids containing single *B* genes (9 and 79) are shown in Fig. 1, cluster 3. The map of cosmid 62 of Deverson *et al.* (1991) which contains matching restriction sites is also included. Although the nucleotide sequences of the second exons of 9 and 79 were

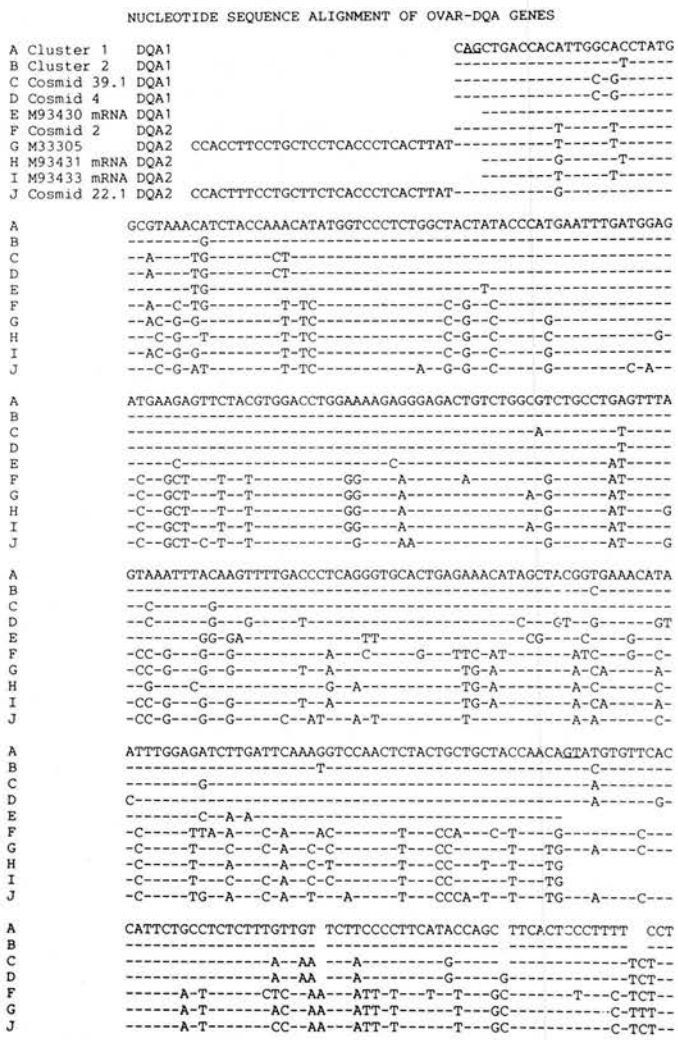


Fig. 2. Multiple nucleotide sequence alignment of the second exons and flanking introns of *Ovar-DQA* genes using the seqnet clustal V alignment program (Higgins *et al.* 1992). A, cluster 1 and EMBL Accession No. M33304 of Scott *et al.* (1991a). B, cluster 2. C, D, F, J cosmids 39.1, 4, 2, and 22.1 of Deverson *et al.* (1991). G, *Ovar-DQA2* of Scott *et al.* (1991a), EMBL Accession No. M33305. E, H, I EMBL Accession Nos M93430, M93431 and M93433 (Lund, unpublished). Identity is represented by –, spaces are inserted to maximize alignment.

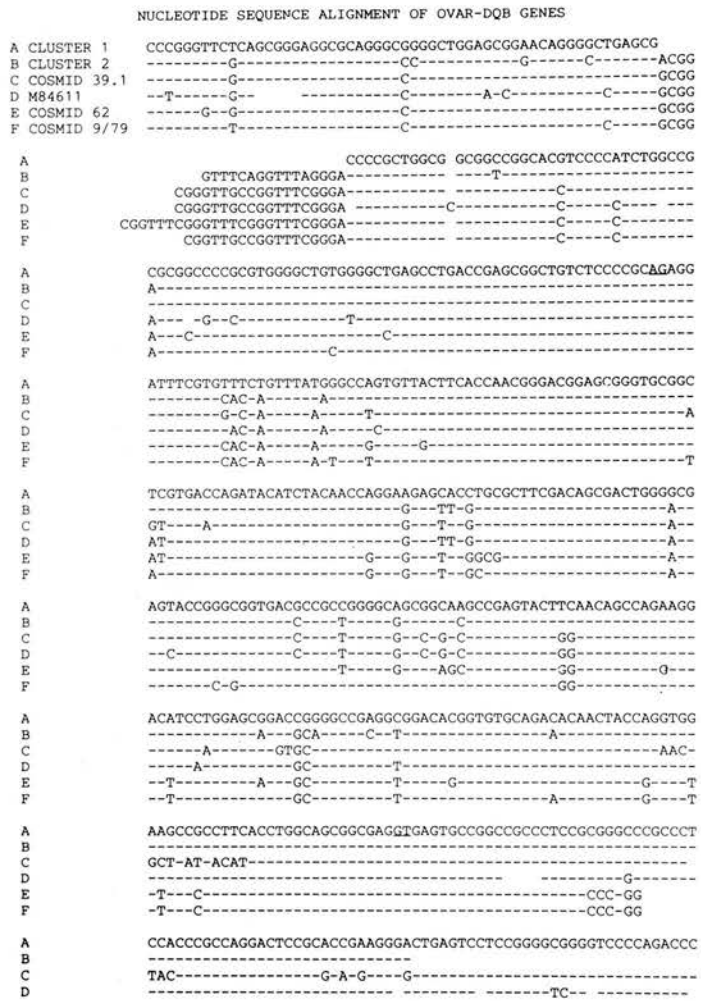


Fig. 3. Multiple nucleotide sequence alignment of the second exons and flanking introns of *Ovar-DQB* genes. A, B, F, *DQB* genes from clusters 1, 2 and 3. C and E, cosmids 39.1 and 62 of Deverson *et al.* (1991). D, *DQB* gene, EMBL Accession No. M84611 (Scott *et al.* 1991b). Identity is represented by –, spaces are inserted to maximize alignment.

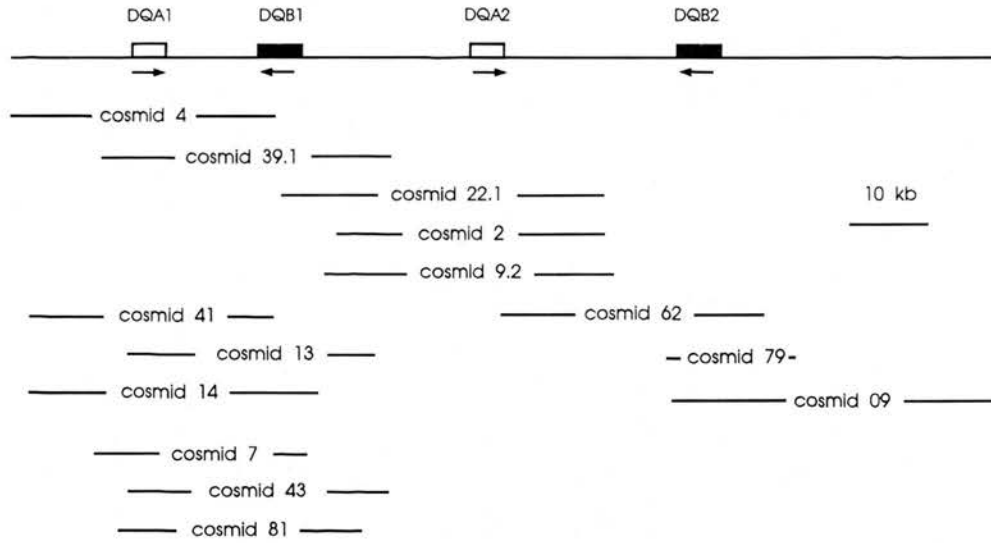


Fig. 4. Map of the *Ovar-DQ* subregion assembled from overlapping cosmid clones. Arrows indicate the direction of transcription of the *Ovar-DQA1*, *DQB1*, *DQA2* and *DQB2* genes. Cosmids in clusters 1, 2 and 3 have been described here. Cosmids 2, 4 and 62 are derived from the pTL6 library described by Deverson *et al.* (1991). Cosmids 22.1 and 39.1 are derived from the pTL5 library described by Deverson *et al.* (1991).

identical (Fig. 3), the differences in the restriction maps 5' of the gene indicated that they represented two alleles at a second *DQB* locus *DQB2*. The sequence of the second exon of the *DQB* gene in cosmid 62 (Fig. 3) was 95% similar to that in cosmids 9 and 79. This similarity, together with the partial identity of the restriction maps indicated that the *B* gene in cosmid 62 was also an allele at this *DQB2* locus. The *B* genes at this locus were 91–92% similar to the *DQB* genes in clusters 1 and 2.

The *A* and *B* genes in the cosmids described by Deverson *et al.* (1991) were given their subtype designation on the basis of their hybridization to HLA class II probes. This has now been confirmed for the *DQ* genes by nucleotide sequencing. The nucleotide sequences of the second exons of the *Ovar-DQA* and *B* genes described by Deverson *et al.* (1991) are compared with each other and with others within the EMBL database in Figs 2 and 3. The sequences of the second exons of the *DQA* genes in cosmids 4 and 39.1 were 92–95% similar to the *DQA1* genes in clusters 1 and 2 described here (Table 1). On the other hand, the *A* genes in cosmids 2/9.2 and 22.1 were only 76% similar to those in clusters 1 and 2 but 87 and 90% similar to the *DQA2* gene of Scott *et al.* (1991a) (Table 1). A *DQA2* gene equivalent to those in cosmids 2/9.2 and 22.1 was not cloned from the pCos8 library. However its presence can be inferred from Southern analysis of DNA from which the library was constructed using an exon 2-specific probe from the *DQA2* gene in cosmid 2 (data not shown). The sequence of the second exons of the *B* genes in overlapping cosmids 39.1 and 22.1 were the same and demonstrated >90% identity to all other ovine-DQB genes described in Fig. 3.

Table 1. Percentage nucleotide similarities between ovine MHC class II DQA genes

Cosmids	DQA1	13	7	4	39.1	DQA2	2/9.2
13	100						
7	98	98					
4	92	92	91				
39.1	95	95	93	95			
DQA2	78	78	78	77	78		
2/9.2	75	75	75	73	77	90	
22.1	79	79	75	75	77	87	88

The percentage similarity between the sequences of exon 2 from the *DQA* genes in the cosmids described here, those described by Deverson *et al.* (1991) and the ovine *DQA1* and *DQA2* genes of Scott *et al.* (1991a). EMBL Accession Nos M33304 and M3305.

A cosmid map of the Ovar-DQ *subregion*

The high sequence similarity between the *A* genes in clusters 1 and 2 described here and those in cosmids 4 and 39.1 of Deverson *et al.* (1991) on the one hand, and the similarity between the *A* genes in cosmids 2/9.2 and 22.1 on the other, prompted us to examine the restriction maps of all *DQ* cosmids with a view to piecing together a map of the *DQ* subregion in the sheep. A schematic map is shown in Fig. 4. This map is made up of 14 cosmids, isolated from three cosmid libraries, produced from three unrelated sheep. The origin of individual cosmids is described in the legend to Fig. 4. At

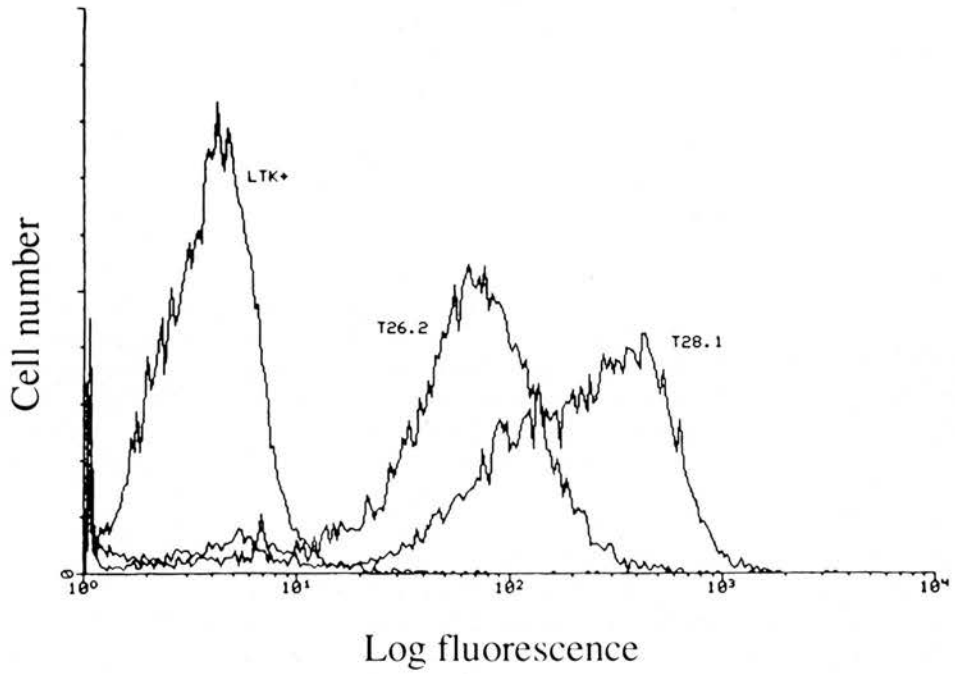


Fig. 5. FACS analysis of mouse L cell lines: Ltk+, T26.2 and T28.1. Sheep class II expression was detected with a cocktail of SBU II monoclonal antibodies (SBU28.1, 37.68, 38.27, 42.40 and 49.1 purchased from M. Brandon, University of Melbourne, Australia) followed by FITC-IgG antimouse Ig as a second-stage reagent (Dakopatts, Clostrup, Donmark). Ltk+, negative control transfected with the thymidine kinase gene only. T26.2 was transfected with cosmid 7 from cluster 2. T28.1 was transfected with cosmid 14 from cluster 1.

least nine restriction sites in the area of overlap between cosmids 39.1 and 22.1 matched with sites at the left-hand side of cosmids 2 and 9.2 of Deverson *et al.* (1991). Thirteen restriction sites at the right-hand side of cosmids 2/9.2 matched up with those on the left side of cosmid 62. As shown in Fig. 1, cosmid 62 matched up with cosmids 9 and 79. This produced a linear tract of 130 kb of DNA linking the expressed *Ovar-DQA1* and *B1* genes to the transcribed but apparently untranslated *Ovar-DQA2* and *DQB2* genes. The *A* and *B* genes were orientated tail-to-tail. The *DQA1* and *DQB1* genes were 11 kb apart while the *DQA2* and *DQB2* genes were 25 kb apart. The distance between the 5' end of the *DQA1* gene and the 5' end of the *DQB2* was 75 kb.

Expression of DQ *class II genes*

Five of the six cosmids in clusters 1 and 2 (Fig. 1) expressed class II molecules on greater than 50% of cells in the primary transfected mouse L cell line. After two rounds of cell sorting, L cell lines expressing the product of each cluster were developed. These lines each expressed class II molecules at the surface of >95% of the cells (Fig. 5). The genes in cosmid 43 failed to express at high levels due to a truncation 65 base pairs 5' of the second exon resulting in the loss of the first exon and all 5' regulatory elements (data not shown). Despite this, class II molecules were detected on 2% of L cells following transfection of cosmid 43.

As previously described (Deverson *et al.* 1991; Ballingall *et al.* 1992), MHC glycoproteins were not detectable following transfection of cosmids containing *DQA2* and *B2* genes.

Discussion

In this study we have identified and expressed two alleles at the *Ovar-DQ* locus equivalent to *HLA-DQ1*. The sequence homology between these *Ovar-DQ1* genes, the genes in cosmids 4, 39.1 and the sequence designated *DQA1* by Scott *et al.* (1991a) suggest that they too are alleles at this locus. The sheep has at least two *DQA* loci which are relatively divergent in terms of their nucleotide sequence (Scott *et al.* 1991a). Deverson *et al.* (1991) described a cluster of three overlapping cosmids 4, 39.1 and 22.1 which contained two linked *DQA* genes separated by a *B* gene. They proposed that the *A* gene in cosmid 22.1 corresponded to the *DQA3* described in some *Bota* class II haplotypes (Andersson 1988). However, sequence data now

suggest that the *A* gene in cosmids 2/9.2, 22.1 and the gene designated *DQA2* by Scott *et al.* (1991a) are alleles at a second *DQA* locus, *DQA2*. By comparison of the restriction maps of cosmids containing *Ovar-DQ* genes, short regions of identity were used to link the various clones together.

It must be emphasized that this map has been constructed from cosmids isolated from three unrelated sheep which were heterozygous at the various loci. The map presented here is similar to that of the *HLA-DQ* subregion but is more extensive than that of the *H-2 I-A* region in which the *A* and *B* genes have not been duplicated. The distance between the expressed *Ovar-DQA1* and *B1* genes at 11 kb was similar to that separating the corresponding *HLA-DQ1* and *Bota-DQ1* genes (Kappes & Strominger 1988; Groenen *et al.* 1990). However, the separation between the *Ovar-DQA2* and *B2* genes, at 25 kb, was much greater than that between the corresponding *HLA* genes, the only other pairing available for comparison. This large separation between *Ovar-DQA2* and *B2* genes explains why cosmids containing both these genes were not identified. The *Ovar-DQ* subregion is more compact than the *HLA-DQ* subregion. The distance between *Ovar-DQB1* and *DQA2* at 22 kb, is considerably less than the 70 kb separating the equivalent *HLA-DQ* genes (Campbell & Trowsdale 1993). No ovine equivalent of the *Bota-DQA3* (Andersson 1988) nor the *HLA-DQB3* pseudogene (Ando *et al.* 1989) has been identified.

The expression of the *DQ1* genes in clusters 1 and 2 contrasts with the consistent failure to obtain expression of the *DQ1* genes in cosmids 4 and 39.1. This may simply have been due to truncation of the *B* gene or its associated regulatory elements. The pattern of transcription of these genes following transient transfection (unpublished data) supports this conclusion. The ovine *DQA2*-specific probe confirmed the results of Scott *et al.* (1991a) in identifying high levels of transcription of *DQA2* genes in peripheral blood mononuclear cells and in macrophages (data not shown). *DQB2* mRNA could also be detected following transient transfection of mouse L cells. The production of *Ovar-DQ2* mRNA both *in vitro* and *in vivo*, suggests that a protein product may be produced. Failure to detect such a product may be due to the lack of a suitable monoclonal antibody.

Experiments involving inoculation of C3H mice with transfected L cells in order to produce polyclonal and monoclonal antisera are under way. Sequential immunoprecipitation and aminoterminal sequence analysis undertaken by Puri *et al.* (1987) identified *Ovar-DQ1* molecules. However, the products of the *DQ2* locus were not identified and remain to be demonstrated. *HLA-DQ* determinants are the products of *HLA-DQA1/B1* genes (Auffray *et al.* 1987). *HLA-DQA2/B2* are not transcribed at high levels in normal B cells, IFNγ-treated vascular endothelial cells or transfected L cells (Collins *et al.* 1984; Auffray *et al.* 1987). The duplication event leading to the presence of two *DQ* loci in man has been suggested to have occurred independently of that in sheep (Fabb *et al.* 1993). Therefore, the failure by Scott *et al.* (1991a) and Fabb *et al.* (1993) to identify a *DQA1* gene in approximately 10% of Merino and Romney sheep tested suggests that any cell surface expression of DQ molecules in these sheep would be the product of a functional *Ovar-DQ2* locus.

The nucleotide sequences of all the *Ovar-DQB1* and *B2* genes described here were >90% similar to one another and hence could not be used to assign a *B* gene to a given locus. Only their proximity to an *Ovar-DQA1* or *Ovar-DQA2* gene could be used to discriminate between these genes.

By combining the data of Scott *et al.* (1987, 1991a,b), Deverson *et al.* (1991) and Fabb *et al.* (1993), with that presented here, the *DQ* subregion of the ovine MHC has been shown to include an expressed *DQA1/B1* locus with *A* and *B* genes separated by approximately 11 kb and orientated in a tail-to-tail manner. The *DQ1* genes are polymorphic. So far five alleles have been identified at the *A1* locus and four alleles at the *B1* locus. We have shown that the *DQ1* locus is linked to a *DQ2* locus with 22 kb separating *DQB1* from *DQA2*. A similar distance separated *DQB2* from *DQA2*. Five alleles have been identified at the *DQA2* locus and three at the *DQB2* locus. These genes are transcribed but cell surface products remain to be demonstrated.

Acknowledgements

We thank A. Sanderson for assistance with cell sorting and D. Haig for help and advice in the preparation of this manuscript. This work benefited from the use of the SEQNET facility. All L cell lines described in this paper are available to all interested groups.

EMBL nucleotide sequence database accession numbers for all sequences presented here are: DQA1 gene in cluster 2, Z28418; DQA2 gene in cosmid 2, Z28419; DQA1 gene in cosmid 4, Z28420; DQA2 gene in cosmid 22.1, Z28421; DQB1 gene in cluster 1, Z28422; DQB1 gene in cluster 2, Z28423; DQB1 gene in cosmid 39.1,

Z28424; DQB2 gene in cosmid 62, Z28425; DQA1 gene in cosmid 39.1, Z28518; and DQB2 gene in cosmid 09, Z28523.

This work was supported by the Scottish Office, Agriculture and Fisheries Department.

References

Andersson L. (1988) Organisation of the bovine MHC class II region as revealed by genomic hybridisations. *Animal Genetics* **19,** 32–4.

Ando A., Kawai J., Maeda M., Tsuji K., Trowsdale J. & Inoko H. (1989) Mapping and nucleotide sequence of a new HLA class II light chain gene, DQB3. *Immunogenetics* **30,** 243–9.

Auffray C., Lillie J.W., Korman A.J., Boss J.M., Frechin N., Guillemot F., Cooper J., Mulligam R.C. & Strominger J.L. (1987) Structure and expression of HLA-DQ alpha and DX alpha genes. *Immunogenetics* **26,** 63–73.

Ballingall K.T., Wright H., Redmond J., Dutia B.M., Hopkins J., Lang J., Deverson E.V., Howard J.C., Puri N. & Haig D. (1992) Characterisation of ovine major histocompatibility complex class II (OLA-DR) genes. *Animal Genetics* **23,** 347–59.

Campbell R.D. & Trowsdale J. (1993) Map of the human MHC. *Immunology Today* **14,** 349–52.

Chardon P., Kirszenbaum M., Cullen P.R., Geffrotin C., Auffray C., Strominger J.L., Cohenn D. & Vaiman M. (1985) Analysis of sheep MHC using class I, II & C4 cDNA probes. *Immunogenetics* **22,** 349–58.

Collins T., Korman A., Wake C., Boss J. & Kappes D. (1984) Immune interferon activates multiple class II major histocompatibility complex genes and the associated invariant chain gene in human endothelial cells and dermal fibroblasts. *Proceedings of the National Academy of Sciences, USA* **81,** 4917–21.

Deverson E.V., Wright H., Watson S., Ballingall K., Huskisson N., Diamond A.G. & Howard J.C. (1991) Class II MHC genes of the sheep. *Animal Genetics* **22,** 211–25.

Ehrlich E., Craig A., Poustka A., Frischauf A.M. & Lehrach H. (1987) A family of cosmid vectors with the multi-copy R6K replication origin. *Gene* **57,** 229–37.

Fabb S.A., Maddox J.F., Goglin-Ewens K.J., Baker L., Wu M.-J. & Brandon M.R. (1993) Isolation, characterisation and evolution of ovine major histocompatibility complex class II DRA and DQA genes. *Animal Genetics* **24,** 249–55.

Germain R.N. & Margulies D.H. (1993) The biochemistry and cell biology of antigen processing and presentation. *Annual Review of Immunology* **11,** 403–50.

Groenen M.A.M., Van Der Poel J.J., Dijkhof R.J.M. & Giphart M.J. (1990) The nucleotide sequence of bovine MHC class II DQB and DRB genes. *Immunogenetics* **31,** 37–44.

Hediger R., Ansari H.A. & Stranzinger G.F. (1991) Chromosome banding and gene locations support extensive conservation of chromosome structure between cattle and sheep. *Cytogenetics and Cellular Genetics* **57,** 127–34.

Herrmann B.G. & Frischauf A.-M. (1987) Isolation of genomic DNA. In: *Methods in Enzymology*, vol. 152. Academic Press, New York.

Higgins D.Q., Bleasby A.J. & Fuchs R. (1992) Clustal-V-improved software for multiple sequence alignment. *Computer Applications in the Biosciences* **8,** 189–91.

Kappes D. & Strominger J.L. (1988) Human class II major histocompatibility complex genes and proteins. *Annual Review of Biochemistry* **57,** 991–1028.

Klein J., Bontrop R.E., Dawkins R.L., Erlich H.A, Gyllensten U.B., Heise E.R., Jones P.P., Parnam P., Wakeland E.K. & Watkins D.I. (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics* **31,** 217–19.

Lang J.C., Wilkie N.M. & Spandidos D.A. (1983) Characterisation of eucaryotic control signals by assay of herpes simplex virus type 1 thymidine kinase. *Journal of General Virology* **64,** 2679–96.

Puri N.K., Walker I.D. & Brandon M.R. (1987) N-terminal amino acid sequence analysis of the A and B polypeptides from four distinct subsets of sheep major histocompatibility complex class II molecules. *Journal of Immunology* **236,** 2996–3002.

Racioppi L., Ronchese F., Schwartz R. & Germain R.N. (1991) The molecular basis of class II MHC allelic control of T cell responses. *Journal of Immunology* **147,** 3718–27.

Sanger F., Nicklen S. & Coulson A.R. (1977) DNA sequencing with the chain terminating inhibitors. *Proceedings of the National Academy of Sciences, USA* **74,** 5463–7.

Scott P.C., Choi C. & Brandon M.R. (1987) Genetic organisation of the ovine MHC class II region. *Immunogenetics* **25,** 116–22.

Scott P.C, Gogolin-Ewens K.J., Adams T.E. & Brandon M.R. (1991a) Nucleotide sequence, polymorphism, and evolution of ovine MHC class II DQA genes. *Immunogenetics* **34,** 69–79.

Scott P.C., Maddox J.F., Gogolin-Ewens K.J. & Brandon M.R. (1991b) The nucleotide sequence and evolution of ovine MHC class II B genes: DQB and DRB. *Immunogenetics* **34,** 80–7.

Steinmetz M.,Winoto A., Minard K. & Hood L. (1982) Clusters of genes encoding mouse transplantation antigens. *Cell* **28,** 489–98.

Zehetner G., Frishauf A. & Lehrach H. (1987) Approach to restriction map determination. *Nucleic Acid and Protein Sequence Analysis: A Practical Approach* (ed. by M.J. Bishop & C.J. Rawlings), pp. 147–64. ILR Press, Oxford.

Animal Genetics, 1995 **26**, 79–84

Analysis of the fine specificities of sheep major histocompatibility complex class II-specific monoclonal antibodies using mouse L-cell transfectants

K T Ballingall*, B M Dutia, J Hopkins, H Wright

Summary

The fine specificities of two panels of monoclonal antibodies (mAbs) for sheep major histocompatibility complex (MHC) class II molecules were determined using five mouse L-cell transfectants, each expressing a defined sheep *DQ* or *DR* MHC class II *A/B* gene pair. Using the transfectants in an indirect fluorescence antibody assay, previous immunochemical characterization of the mAbs was confirmed for 16 of 23 mAbs tested. The MHC class II subtype specificity (*DQ* or *DR*) of each mAb was assigned without interference from the products of other expressed class II loci. This allowed the identification of both cross-locus specificities as well as defining fine specificities of mAbs previously only partially characterized by immunochemical techniques.

Keywords: MHC class II, monoclonal antibodies, sheep, transfection

Introduction

The class II genes within the major histocompatibility complex (MHC) of all vertebrate species studied encode highly polymorphic cell surface glycoproteins that present processed fragments of foreign antigen to $CD4^+$ helper T-lymphocytes (Klein 1986; Braciale *et al.* 1987). The high levels of polymorphism, associated with many of the MHC class II genes within an outbred population, are associated with variations in individual immune responses to antigen and susceptibility or resistance to some autoimmune diseases (Sinha *et al.* 1989).

The ovine MHC (*Ovar-Mhc*) contains class II loci corresponding to all of those found in the human (*HLA*) complex with the exception of the *HLA-DP* locus (Chardon *et al.* 1985; Scott *et al.* 1987, 1991a,b; Deverson *et al.* 1991; Ballingall *et al.* 1992; Wright & Ballingall 1994). In addition, the ovine (Wright *et al.* 1994), bovine (van der Poel *et al.* 1990) and caprine (Mann *et al.* 1993) MHC each contain a locus designated *DY*. Nucleic acid probe and oligonucleotide reagents have allowed the analysis of individual sheep MHC class II loci either by restriction fragment length polymorphism (RFLP) (Grain *et al.* 1993) or polymerase chain reaction (PCR) (Schwaiger *et al.* 1993). However, to study the distribution and function of sheep MHC class II proteins during the immune response, mAbs specific for the products of each of the expressed sheep MHC class II loci are required. Previously, immunochemical and protein sequencing techniques have been used to analyse the specificity of sheep MHC class II specific mAbs (Puri *et al.* 1987; Puri & Brandon 1987; Hopkins *et al.* 1986; Dutia *et al.* 1990, 1993, 1994). In this analysis two panels of MHC class II-specific mAbs were characterized using mouse L-cell transfectants, each of which expresses a single MHC class II molecular type (*DQ* or *DR*). In a previous publication, mAbs specific for a *DR* transfectant were identified (Ballingall *et al.* 1992). This paper aims to expand on that study by characterizing the specificity of the mAbs for L-cell lines expressing two alleles at the *DQ1* locus (Wright & Ballingall 1994) plus an additional L-cell line expressing an allele at the *DR1* locus.

Materials and methods

Sheep MHC class II genes

The construction and screening of genomic cosmid libraries and restriction mapping of cosmid clones containing sheep MHC class II genes have been described previously (Deverson *et al.* 1991; Wright & Ballingall 1994). The second exons of MHC class II genes were subcloned into M13mp18/19 and sequenced by the dideoxy chain termination method (Sanger *et al.* 1977) using Sequenase 2·0 or *Taq*uence enzymes together with the 7-deaza dGTP analogue (United States Biologicals, Cleveland, OH) to overcome problems with compressions.

K T Ballingall
H Wright
Moredun Research Institute, 408 Gilmerton Road, Edinburgh EH17 7JH, UK
B M Dutia
J Hopkins
Department of Veterinary Pathology, University of Edinburgh, Summerhall, Edinburgh EH9 1QH, UK

*Present address: ILRAD, PO Box 30709, Nairobi, Kenya.
Correspondence: K T Ballingall.

Accepted 8 November 1994

Mouse L-cell transfectants

The generation of mouse L-cell lines expressing the products of transfected sheep MHC class II genes at the cell surface has been described previously (Ballingall *et al.* 1992). Briefly, 6 μg of cosmid DNA was transfected into mouse Ltk$^-$ cells along with the herpes simplex virus thymidine kinase gene (HSVtk, Lang *et al.* 1983) as a selection marker. Following HAT selection, L-cells expressing the MHC class II molecules at the cell surface were identified by an indirect fluorescence antibody (IFA) test using a cocktail of MHC class II-specific mAbs (SBUII 28.1, 37.68, 38.27, 42.2 and 49.1) as a primary reagent and FITC-conjugated anti-mouse immunoglobulin as the second labelled antibody. Cells were analysed on the FACSCAN and sorted using the FACS STAR (Becton Dickinson, Sunnyvale, California, USA). All L-cell transfectants were cultured in Dulbecco's modified Eagles medium (DMEM, Gibco-BRL, Paisley, UK) supplemented with 10% heat inactivated fetal bovine serum, penicillin (60 μg/ml), streptomycin (100 μg/ml). Cells were harvested prior to confluence for analysis.

Table 1. Ovine MHC class II-specific monoclonal antibodies used in this study

Monoclonal antibody	Ig isotype	Alpha/Beta chain specificity	Previously proposed subtype specificity
SBUII 28.1, 37.68	IgG1	Conformational	DR
SBUII 38.27	IgG1	Conformational	DQ
SBUII 42.2	IgG1	Alpha	DR
SBUII 49.1	IgG2a	Beta	Pan-beta
SW73.2 (Rat)	IgG2b	Beta	Pan-beta
VPM 1	IgM	Conformational	ND
VPM 4, 46	IgG2a	Beta	ND
VPM 45	IgG2a	Beta	DQ
VPM 36	IgG1	Alpha	DQ
VPM 38, 47, 54	IgG1	Alpha	DR
VPM 58, 59	IgG1	Conformational	DR
VPM 37, 57, 43	IgG1	Beta	DR
VPM 41	IgG1	Beta	DQ
VPM 16, 40, 44	IgG1	Beta	DQ

The SBUII mAbs have been described previously (Puri *et al.* 1985, 1987; Puri & Brandon 1987). Their proposed specificity was based on N-terminal sequence analysis. The VPM mAbs have also been described previously (Hopkins *et al.* 1986; Dutia *et al.* 1990, 1993, 1994). The subtype specificity of VPM 36, 37, 38 and 54 was determined by N-terminal sequence analysis (Dutia *et al.* 1993, 1994). The specificity of the remainder of the VPM mAbs was determined by ELISA of mAb supernatants against affinity purified ovine *DQ* and *DR* antigens. (Dutia *et al.* 1993, 1994; B. M. Dutia, unpubl. obs.). SW73.2 reacts with both *DQ* and *DR* antigens (Hopkins *et al.* 1986; Dutia *et al.* 1993). ND indicates that subtype specificity has not been determined.

Monoclonal antibodies

The sheep MHC class II-specific mAbs used in this study are listed in Table 1.

Screening of mAbs using L-cell transfectants

The specificity of each mAb for each L-cell line was determined by the IFA assay. Cells were analysed by FACSCAN and positive antibodies were identified as those that increased FL-1 FITC (green) fluorescence relative to cells transfected with the HSVtk gene only (Ltk$^+$ control).

Results

Mouse L-cell transfectants

Two *DQ* transfectant cell lines, DQ-T28.1 and DQ-T26.2, expressed the products of the polymorphic *DQA1* and *B1* genes described by Wright & Ballingall (1994). These represented two alleles at the *Ovar-DQ1* locus, isolated from a heterozygous sheep. A third *DQ* transfectant, DQ-T37.3, expressed a product which was derived from an *A/B* gene pair in which the *DQA1* gene was truncated as described by Wright & Ballingall (1994).

The generation of the L-cell transfectant, DR-T8.1, expressing an *Ovar-DR* protein was described by Ballingall *et al.* (1992). A new *DR* transfectant, DR-T31.3, was produced by cotransfecting L-cells with the expressed *Ovar-DRA* gene (Ballingall *et al.* 1992) together with cosmid 02, which contained an *Ovar-DRB1* gene isolated from the library described by Wright & Ballingall (1994). The nucleotide sequence of the second exon of the *DRB1* gene is shown in Fig. 1. The sequence was 95% similar to the expressed *Ovar-DRB1* gene described by Ballingall *et al.* (1992). The predicted amino acid sequence of these genes differed by eight amino acids (Fig. 1).

Screening of mAbs for Ovar-DQ and DR specificity

Each transfected L-cell line was used to screen the two panels of mAbs for cell surface *DQ* and *DR* specificity. The specificity of each mAb for the MHC class II molecule expressed on the surface of each transfectant is summarized in Table 2. The sheep MHC class II-specific mAbs that recognized all transfectants expressing intact *DQ* and *DR* molecules at the cell surface were VPM 38, 46 and SW73.2. Monoclonal antibodies SBUII 28.1, 42.2 and 49.1 each recognize a determinant found on three of the four lines. Of these mAbs, only SW73.2 reacted

A

```
                    exon 2
Ovar-DRB1   ctcgagCACATTTCTTGGAGTATACTAAGAAAGAGTGTCGTTTCTCCAACGGGACGGAGCGGGTGCGGTTCCTGGACAGATACTTCTATAATGGAGAAGA
Cosmid 02   --gc------------------CA-----GC-------------------------------------A----------------------------

Ovar-DRB1   GTACGTGCGCTTCGACAGCGACTGGGGCGAGTACCGAGCGGTGGCCGAGCTGGGGCGGCCGGACGCCAAGTACTGGAACAGCCAGAAGGAGATCCTGGAG
Cosmid 02   ----------------A-------------------------------------------------------------------------CT--------

Ovar-DRB1   CGGAGGCGGACCGAGGTGGACACGTACTGCAGACACAACTACGGGGTCATTGAGAGTTTCACTGTGCAGCCGCCAGgtgagcgcgggggtgggcggccag
Cosmid 02   -A--C-------------------------------------------------------------------------------------a---------

Ovar-DRB1   tgtggagcagtgtgtgtgcgtgtgtgcgtgtgcgtgtgcgtgtgcgtgtgcgtgtgtgtgtgtgtgtgtgtgtg
Cosmid 02   ---------..-----------------------------------t-----t----------a-a-a-a-a-a-a-
```

B

```
                    exon 2
Ovar-DRB1   HFLEYTKKECRFSNGTERVRFLDRYFYNGEEYVRFDSDWGEYRAVAELGRPDAKYWNSQKEILERRRTEVDTYCRHNYGVIESFSVQRR
Cosmid 02   -----H-S----------Y---------------N----------------------DF--QT----------------------
```

Fig. 1.A. Comparison of the nucleotide sequence of exon 2 of the *Ovar-DRB1* gene described by Ballingall *et al.* (1992) with that of the new allele at the expressed locus described here. Coding nucleotide sequences are in upper case. Dashes indicate identity and gaps introduced into intronic sequences to maintain alignment are indicated by dots. The splice junctions are underlined. **B:** Comparison of the predicted amino acid translations of exon 2 of the genes described above. Dashes indicate identity.

strongly with the transfectant expressing the truncated *DQA* gene (DQ-T37.3). VPM 38, 46 and SBU 42.20 failed to recognize this transfectant and SBUII 28.1 reacted, but with a much reduced fluorescence intensity compared to its reactivity with the transfectant expressing the products of the complete *DQA* and *B* genes (DQ-T26.2).

Table 2. The specificity of the panel of monoclonal antibodies for the *DQ*- and *DR*-expressing L-cell lines.

Monoclonal antibody	Transfectant DQ-T28.1	Transfectant DQ-T26.2	Transfectant DR-T8.1	Transfectant DR-T31.3
SW73.2	+++	+++	+++	+++
SBUII 28.1	+++	+++ (+)	−	+++
SBUII 42.2	−	+++ (−)	+++	+++
SBUII 49.1	+++	−	+++	+++
VPM 38	++	++ (−)	+++	+++
VPM 46	++	++ (−)	+++	+++
SBUII 38.27	+++	+++	−	−
VPM 1	++	+++	−	−
VPM 4	+++	+++	−	−
VPM 16	+++	+++ (−)	−	−
VPM 36	+++	+++	−	−
VPM 40	−	+++	−	−
VPM 41	−	+++	−	−
VPM 44	−	+++	−	−
VPM 45	−	+++ (−)	−	−
VPM 47	−	−	+++	+
SBUII 37.68	−	−	+++	+++
VPM 54	−	−	+++	+++
VPM 57	−	−	+++	+++
VPM 58	−	−	++	++
VPM 59	−	−	++	++
VPM 37	−	−	−	+
VPM 43	−	−	−	++

Binding of individual mAbs to the transfected L-cell line relative to the tk^+ L-cell control: +++, strong; ++, variable; +, week; −, no binding. Differences in the binding of the mAbs to the truncated *DQ* L-cell line DQ-T37.3, when compared to the complete transfectant DQ-T26.2, are indicated in parenthesis.

Ovar-DQ1 subtype specificity was assigned to mAbs SBUII 38.27 and VPM 1, 4, 16 and 36, as each recognized both *DQ* transfectants. In contrast, the mAbs VPM 40, 41, 44 and 45 were able to differentiate between the proteins expressed on the two *DQ* transfectant cell lines DQ-T26.2 and DQ-T28.1. Each of these mAbs recognized transfectant DQ-T26.2 and was therefore considered haplotype-specific. None of the mAbs was specific for only transfectant DQ-T28.1. Only SBUII 49.1, which reacts with a broad range of sheep MHC class II beta chains, recognized this *DQ* transfectant and failed to recognize DQ-T26.2. However, SBUII 49.1 also recognized both *DR* transfectants. Of the *DQ*-specific mAbs, only VPM 16 and 45 failed to recognize the truncated transfectant DQ-T37.2.

The mAbs SBUII 37.68, VPM 54, 57, 58 and 59 reacted with both *DR* transfectants and were considered *DR* subtype specific. *DRB1* allele specificity (i.e. mAbs that distinguished between the *DRB1* gene products) was assigned to mAbs VPM 37 and 43. SBUII 28.1 recognized the *DR* line DR-T31.3 but failed to recognize the DR-T8.1 line. However, this mAb also recognized both *DQ* L-cell lines. SBUII 42.20 recognized both *DR* lines and also a determinant present on the *DQ* line DQ-T26.2.

Discussion

Transfected L-cells expressing a single subtype of MHC class II molecule have been previously used to generate novel mAbs and to assign MHC class II subtype, locus and allele specificity to panels of mAbs directed against HLA (Tosi *et al.* 1986; Klohe *et al.* 1988), and H-2 determinants

(Landais *et al.* 1986; Lechler *et al.* 1986). The generation of a large panel of transfected L-cell lines expressing HLA class II molecules and their use in the production of novel mAb reagents also formed a major part of the 11th International Histocompatibility Workshop (Inoko *et al.* 1991). This report describes the application of these techniques to the characterization of mAbs specific for products of the *Ovar-Mhc* class II region. L-cells expressing sheep MHC class II molecules have enabled the fine specificity of two panels of sheep MHC class II-specific mAbs to be determined. By using a panel of transfectants we were able to identify 'holes' or 'gaps' in the specificity of mAbs previously thought to recognize a monomorphic determinant found on all sheep class II molecules. For example, SBUII 49.1, previously considered to be a pan-beta chain-specific mAb (Puri *et al.* 1987), failed to recognize one of the *DQ* L-cell transfectants. One possible explanation for this result is that the MHC class II molecules expressed on the surface of mouse L-cells differ from those found on sheep cells *in vivo*. However, SBUII 49.1 recognized the *DQ* and *DR* beta proteins expressed on all the other mouse L-cell transfectants. Only mAbs SW73.2 and VPM 46 recognized the beta chains expressed on all the transfectant cell lines. The panel of transfectant cell lines also allowed the identification of cross-locus specificity, i.e. those mAbs that recognize the products of both *DQ* and *DR* loci. For example, by immunochemical and N-terminal amino acid sequencing techniques VPM 38 was previously considered *DRA*-specific (Dutia *et al.* 1990, 1993), but by using *DR* and *DQ* transfectants we have demonstrated that VPM 38 also reacted with *DQ* molecules at the surface of mouse L-cells. However, when the class II molecules were isolated from the transfected cell lines, separated on SDS-PAGE and blotted, the reactivity of VPM 38 was purely *DR*-alpha chain-specific (B. Dutia, personal communication). In view of these data, VPM 54 is the *DRA*-specific mAb of choice for both cell surface labelling and immunoblotting techniques. Monoclonal antibody SBUII 28.1 also recognized both *DQ* transfectants. This mAb was previously considered, by sequential immunoprecipitation and N-terminal sequence analysis, to be *DR*-specific (Puri *et al.* 1987; Puri & Brandon 1987). On the other hand, the specificity of SBUII 37.68 and 38.27 for monomorphic conformational determinants present on *DR* and *DQ* molecules respectively (Puri *et al.* 1987) was confirmed using the transfectants. This was also the case for other reagents such as VPM 36, a monomorphic *DQ* alpha chain-specific mAb (Dutia *et al.* 1993). The *DR* alpha and *DR* beta chain specificities of VPM 54 and VPM 57 respectively were also confirmed, as these mAbs recognized both *DR* transfectants but failed to recognize either *DQ* line. The mAbs that recognize polymorphic determinants of *DQ* molecules (VPM 40, 41, 44, and 45) are all beta-chain specific, and each recognizes the *DQB* gene product in DQ-T26.2 but not in DQ-T28.1. The production of a larger panel of *DQ* transfectants would allow the fine specificity of the polymorphic MHC class II-specific mAbs to be further characterized. Monoclonal antibodies VPM 1 and 4 had not been previously assigned a subtype specificity, but could be assigned to *DQ* on the basis of this study. Similarly, VPM 46 appears to recognize a monomorphic determinant on both *DQ* and *DR* molecules.

The L-cell line DQ-T37.3 is unusual as the MHC class II molecules are expressed at the cell surface, despite the loss of exon 1 and all 5′ regulatory elements of the transfected *DQA* gene. Exon 1 encodes the signal/leader peptide and the first four N-terminal amino acids of the mature membrane-associated glycoprotein (Andersson *et al.* 1987). As the function of the signal/leader peptide is to assist in the transfer of the growing peptide chain from the ribosome into the endoplasmic reticulum (Klein 1986), its loss may be expected to alter the glycosylation and folding of the newly translated protein chain and, therefore, the binding of mAbs. The L-cell transfectant expressing the product of the truncated *DQA* gene altered the binding of six mAbs (SBU 28.1 and 42.20, VPM 16, 38, 45 and 46). However, VPM 36, the only *DQ*-alpha-specific mAb in the test pool, bound with equal affinity to both the complete and truncated transfectants. This indicated that its determinant was not altered by the loss of the first four amino acids of the mature *DQ* alpha chain, or by any alterations resulting from aberrant glycosylation or processing as a result of the loss of the signal/leader peptide. It is of interest that not all of these mAbs are specific for the alpha chain. SBU 28.1 has been reported to recognize a conformational determinant between alpha and beta chains (Puri *et al.* 1987). This determinant must be altered by the 5′ truncation of the *DQA1* gene. VPM 16, 45 and 46 are beta-chain specific, but each loses *DQ* specificity as a result of the truncation of the *DQA* gene. The conformation of the beta chain and therefore the epitopes recognized by VPM 16, 45 and 46 must be influenced by the corresponding alpha chain.

Acknowledgements

Computer analysis of nucleotide sequences made use of the SEQNET molecular biology

computer facility. The EMBL GenBank accession number for the *Ovar-DRB1* sequence presented here is X76905. This work has been supported by funds from the Scottish Office, Agriculture and Fisheries Department.

References

Andersson G., Larhammar D., Widmark E., Servenius B., Peterson P.A. & Rask L. (1987) Class II genes of the human major histocompatibility complex. *Journal of Biological Chemistry* **262,** 8748–55.

Ballingall K.T., Wright H., Redmond J. *et al.* (1992) Characterization of ovine major histocompatibility complex class II (OLA-DR) genes. *Animal Genetics* **23,** 347–59.

Braciale T.J., Morrison L.A., Sweetser M.T., Sambrook J., Gething M.J. & Braciale V.L. (1987) Antigen presentation pathways to class I and class II MHC restricted T lymphocytes. *Immunology Reviews* **98,** 95–114.

Chardon P., Kirszenbaum M., Cullen P.R. *et al.* (1985) Analysis of sheep MHC class I, II & C4 cDNA probes. *Immunogenetics* **22,** 349–58.

Deverson E.V., Wright H., Watson S. *et al.* (1991) Class II MHC genes of the sheep. *Animal Genetics* **22,** 211–25.

Dutia B.M., Hopkins J., Allington M.P., Bujdoso R. & McConnell I. (1990) Characterization of monoclonal antibodies specific for alpha- and beta-chains of sheep MHC class II. *Immunology* **70,** 27–32.

Dutia B.M., McConnell I., Bird K., Keating P. & Hopkins J. (1993) Patterns of major histocompatibility complex class II expression on T-cell subsets in different immunological compartments. 1. Expression on resting T cells. *European Journal of Immunology* **23,** 2882–88.

Dutia B.M., McConnell I., Ballingall K.T., Keating P. & Hopkins J. (1994) Evidence for the expression of two distinct MHC class II *DRB*-like molecules in the sheep. *Animal Genetics* **24,** 235–41.

Grain F., Nain M.-C., Labonne M.-P. *et al.* (1993) Restriction fragment length polymorphism of *DQB* and *DRB* class II genes of the ovine major histocompatibility complex. *Animal Genetics* **24,** 277–84.

Hopkins J., Dutia B.M. & McConnell I. (1986) Monoclonal antibodies to sheep lymphocytes. *Immunology* **59,** 433–8.

Inoko H., Bodmer J.G., Heyes J.M., Drover S., Trowsdale J. & Marshall W.H. (1991) Joint report on the transfectant/monoclonal antibody component. In: HLA 1991, *Volume 1* (ed. by K. Tsuji, M. Aizawa & T. Sasazuki), pp.919–29. Oxford University Press, Oxford.

Klein J. (1986) *The Natural History of the Major Histocompatibility Complex,* John Wiley, New York.

Kloe E.P., Watts R. & Bahl M. (1988) Analysis of the molecular specificities of anti-class II monoclonal antibodies by using L-cell transfectants expressing HLA class II molecules. *Journal of Immunology* **141,** 2158–64.

Landais D., Beck B.N., Buerstedde J.-M. *et al.* (1986) The assignment of chain specificities for anti-Ia monoclonal antibodies using L-cell transfectants. *Journal of Immunology* **137,** 3002–5.

Lang J.C., Wilkie N.M. & Spandidos D.A. (1983) Characterization of eucaryotic control signals by assay of herpes simplex virus type 1 thymidine kinase. *Journal of General Virology* **64,** 2679–96.

Lechler R.I., Ronchese F., Braunstein N.S. & Germain R.N. (1986) I-A restricted T-cell antigen recognition. Analysis of the roles of A-alpha and A-beta using DNA-mediated gene transfer. *Journal of Experimental Medicine* **163,** 678–82.

Mann A.J., Abraham L.J., Cameron P.U., Robinson W., Giphart M.J. & Dawkins R.L. (1993) The caprine MHC contains DYA genes. *Immunogenetics* **37,** 292–5.

van der Poel J.J., Groenen M.A.M., Dijkhof R.J.M., Ruyter D. & Giphart M.J. (1990) The nucleotide sequence of the bovine MHC class II alpha genes: *DRA, DQA* and *DYA*. *Immunogenetics* **31,** 29–36.

Puri N.K., Mackay C.R. & Brandon M.R. (1985) Sheep lymphocyte antigens (OLA). *Immunology* **56,** 725–32.

Puri N.K., Walker I.D. & Brandon M.R. (1987) N-terminal amino acid sequence analysis of the A and B polypeptides from four distinct subsets of sheep major histocompatibility complex class II molecules. *Journal of Immunology* **236,** 2996–3002.

Puri N.K. & Brandon M.R. (1987) Sheep MHC class II molecules. Identification and characterization of four distinct subsets of sheep MHC class II molecules. *Immunology* **62,** 575–80.

Sanger F., Nicklen S. & Coulson A.R. (1977) DNA sequencing with the chain terminating inhibitors. *Proceedings of the National Academy of Sciences of the USA* **74,** 5463–7.

Schwaiger F.-W., Weyers E., Epplen C. *et al.* (1993) The paradox of MHC-*DRB* exon/intron evolution: alpha-helix and beta-sheet encoding regions diverge while hypervariable intronic simple repeats coevolve with beta-sheet codons. *Journal of Molecular Evolution* **37,** 260–72.

Scott P.C., Choi C. & Brandon M.R. (1987) Genetic organisation of the ovine MHC class II region. *Immunogenetics* **25,** 116–22.

Scott P.C., Gogolin-Ewens K.J., Adams T.E. & Brandon M.R. (1991a) Nucleotide sequence, polymorphism, and evolution of ovine MHC class II *DQA* genes. *Immunogenetics* **34,** 69–79.

Scott P.C., Maddox J.F., Gogolin-Ewens K.J. & Brandon M.R. (1991b) The nucleotide sequence and evolution of ovine MHC class II B genes: *DQB* and *DRB*. *Immunogenetics* **34,** 80–7.

Sinha A.A., Acha-Orbea H., Todd J., Timmerman L. & McDevitt H.O. (1990) Functional importance of MHC class II polymorphism in normal immune responses and autoimmune disease. In: *Molecular Biology of HLA Class II Antigens* (ed. by J. Silver), pp.147–68. CRC Press, Boca Raton, Florida.

Tosi R., Tanigaki N., Preval C., Gorski J. & Mach B. (1986) Immunochemical analysis of a cell transfected with an HLA-DR gene reveals a new alloantigenic specificity within HLA-DRw52. *European Journal of Immunology* **16**, 1603–08.

Wright H. & Ballingall K.T. (1994) Mapping and characterization of the *DQ* sub-region of the ovine MHC. *Animal Genetics* **25**, 243–9.

Wright H., Ballingall K.T. & Redmond J. (1994) The *DY* sub-region of the sheep MHC contains an A/B gene pair. *Immunogenetics* **40**, 230–4.

mmunogenetics (1995) 41: 131–133

BRIEF COMMUNICATION

Harry Wright · James Redmond · Frank Wright
Keith T. Ballingall

The nucleotide sequence of the sheep MHC class II *DNA* gene

Received: 15 August 1994 / Revised: 11 October 1994

The human MHC class II *DNA* (formerly *DZα*) gene was identified and sequenced by Trowsdale and Kelly (1985). When a molecular map of the *HLA-D* region became available (reviewed by Campbell and Trowsdale 1993), it was shown that the *HLA-DNA* gene was unusual in not having a *B* gene partner situated within a few kilobases (kb), the nearest *B* gene being *HLA-DPB1*. The nearest unpaired *B* gene is *HLA-DOB* (Tonnelle et al. 1985) which is approximately 160 kb telomeric of *HLA-DNA*. More recently, the mouse MHC class II genes *H-2OA* and *H-2OB* were shown to be equivalent to the *HLA-DNA* and *HLA-DOB* genes (Karlsson et al. 1991; Karlsson and Peterson 1992). Moreover, the mouse genes expressed an MHC class II protein whose tissue distribution was restricted to B cells and epithelial cells of the thymic medulla. No corresponding HLA-DN protein has been reported.

Although the presence of the equivalent genes in other species, including cattle and sheep, has been inferred from Southern analysis of genomic DNA (Scott et al. 1987; Andersson and Rask 1988; Trowsdale et al. 1989), the only other *DNA*-like sequence in the database is that of the tammar wallaby, *Macropus eugenii* (Slade et al. 1994). Deverson and co-workers (1991) described cosmid clones from the sheep MHC class II region which, on the basis of differential hybridization to an *HLA-DNA* probe, appeared to contain the *Ovar-DNA* gene. Within 15 kb of the *A* gene was a restriction fragment which hybridized to class II *B* probes at low stringency. However, subsequent sequencing showed that the putative *B* gene was in fact a region of highly repetitive DNA very closely related to the short interspersed nuclear element (SINE) Bov-A2 described by Lenstra and co-workers (1993). The nucleotide sequence of the *A* gene together with its predicted amino acid translation is shown in Figure 1. It has all the salient features of a class II *A* gene. These include two exons coding for the two extracellular domains, and one coding for a proline-rich connecting peptide, a hydrophobic transmembrane region, and a cytoplasmic tail. It also has the two conserved N-linked glycosylation sites NGT and NAT, indicated in italics in Figure 1, and the two conserved cysteine residues forming the disulphide bond in the alpha-2 domain. The sheep and human genes shared 83% nucleotide identity at exons 2 and 3. This figure agrees well with the 82% identity for the human vs cattle comparison (Andersson and Gustafsson, quoted as a personal communication in Rosen-Bronson and Long 1991). The predicted amino acid translations of exons 2–4 of the published *DNA* genes are shown in Figure 2. The *Ovar-DNA* sequence was most like that of the *HLA-DNA* gene with the amino acid identity at the second exon being 78%. The identity fell to 76%, 53%, 54%, 42%, and 53% when the sheep gene was compared with *H2-OA* and the *Ovar-DQA1*, *DQA2*, *DRA*, and *DYA* second exons, respectively. The sheep gene shared 54% amino acid identity with the tammar wallaby *DNA* gene at exon 2 and 66% at exon 3.

A phylogenetic tree of translated second exons from representative human, mouse, cattle, rabbit, pig, and sheep *A* genes, together with the four *DNA* sequences is shown in Figure 3. The tree was constructed by applying the neighbor-joining method (Saitou and Nei 1987) to pairwise distances that had been corrected for multiple hits in the amino acid sequence by the method of Kimura (1983). Statistical testing of the tree topology was done by using the bootstrap method based on 2000 trials (Hedges 1992). The bootstrap results are summarized as the percentage of trials supporting a given node. For example, the [*HLA-DNA* (*H-2OA*, *Ovar-DNA*)] cluster was supported by 99.8% of trials, suggesting that this cluster was indeed distinct from the rest of the tree. Only bootstrap percentages of over 95% are reported. The PHYLIP package was used for all phylogenetic analyses.

The nucleotide sequence data reported in this paper have been submitted to the EMBL nucleotide sequence database and have been assigned the accession number Z29533

H. Wright (✉) · J. Redmond · K. T. Ballingall
Moredun Research Institute, 408 Gilmerton Road, Edinburgh, EH17 7JH, UK

F. Wright
SASS, University of Edinburgh, J. C. M. B., King's Buildings, Edinburgh, EH9 3JZ, UK

Fig. 1 The nucleotide and amino acid sequences of the *Ovar-DNA* gene. Exon sequences are in *capital letters* with the single-letter amino acid code above each codon. Splice junctions are *underlined* and the putative polyadenylation signal is in *bold type*

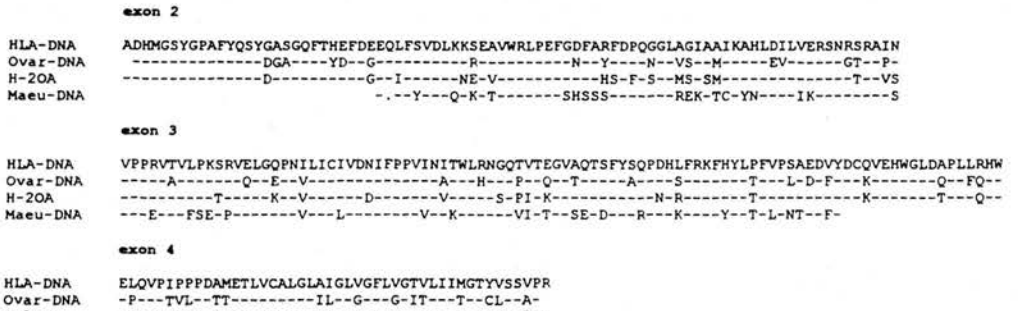


Fig. 2 Multiple alignment of the amino acid translations of exons 2, 3, and 4 of the human, sheep, mouse, and tammar wallaby (*Macropus eugenii*) *DNA* genes

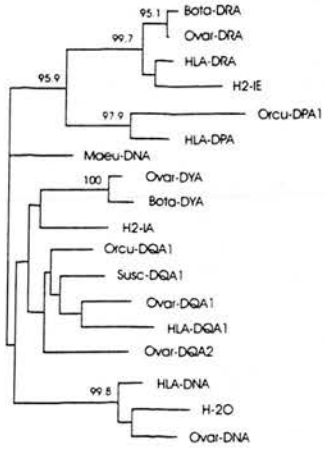


Fig. 3 Phylogenetic tree of MHC class II *A* genes. Tree construction is described in the text. *Ovar*: sheep (*DQA1* and 2, Acc. Nos. M93430 and M93431; *DRA*, Acc. No. Z11600; *DYA*, Acc. No. Z27398; *DNA*, Acc. No. Z29533). *Bota*: cattle (*DYA*, Acc. No. M30119; *DRA*, Acc. No. M30120). *H2*: mouse (*OA*, Acc. No. M95514; *I-E*, Acc. No. J00396; I-A, Acc. No. K01923). *HLA*: **human** (*DNA*, Acc. No. X02882; *DPA1*, Acc. No. X03100; *DRA*, Acc. No. J00204; and *DQA1* Acc. No. M17237). *Orcu*: rabbit (DQA1, Acc. No. M15557; *DPA1*, Acc. No. M22640), *Maeu*: tammar wallaby (*DNA*, Acc. No. 112121) and *Susc*: pig (*DQA1*, Acc. No. M29938)

One feature of the human and mouse genes was the occurrence of a point mutation in the polyadenylation signal producing the sequence ACTAAA, instead of the eukaryotic consensus AATAAA. The RNA transcript of the *HLA-DNA* gene was unusually long at 3.5 kb, and this was ascribed to the altered polyadenylation signal (Trowsdale and Kelly 1985). Subsequently, human and mouse cDNA clones of the correct size were sequenced in which a poly(A) tail was added a few base pairs downstream of the ACTAAA signal (Jonsson and Rask 1989; Young and Trowsdale 1990; Karlsson and Peterson 1992). Rosen-Bronson and Long (1991) sequenced a 3.5 kb cDNA clone derived from the *HLA-DNA* gene and showed that in this case a poly(A) tail was added after a second mutated polyadenylation signal, ATTAAA, 2.3 kb downstream from the first.

A putative polyadenylation signal in the sheep gene is shown in bold type at position 1823 in Figure 1. The sequence AGTAAA carries a different point mutation from that in the human and the mouse. This sequence is approximately one-third as active in polyadenylation as the consensus sequence and about twice as active as the ACTAAA form found in the human and mouse (reviewed by Wickens 1990). To assess whether the size of the transcript from the sheep gene was influenced by the altered signal, northern blots of total RNA prepared from various sheep tissues, including thymus and spleen, were hybridized with an exon 2-specific *Ovar-DNA* probe and with an *Ovar-DRA* probe as control. The *Ovar-DRA* gene had been shown to have the consensus AATAAA signal

ngall et al. 1992). The major band on both blots was er than the 18S rRNA which runs at 2 kb. The ntration of *Ovar-DNA* RNA was at least an order of itude less than that of *DRA* (data not shown). It would r that the sheep transcript is of normal size but of low lance.

he lack of expression of the mouse H2-O protein on ssional antigen-presenting cells, together with the nce of equivalent genes in the diverse species exam- led Karlsson and co-workers (1991) and Karlsson and son (1992) to conclude that these genes have remained ional throughout evolution and to speculate that their ion may be different from that of the other class II cules. Interestingly, Slade and co-workers (1994) ted the presence of two *DNA* loci in the marsupial nar wallaby, only one of which was expressed.

owledgments The clones are freely available to interested par- This work benefitted from the use of the SEQNET facility. The rs are supported by funds from the Scottish Office Agriculture isheries Department.

rences

rsson, L. and Rask, L. Characterisation of the MHC class II gion in cattle. The number of *DQ* genes varies between haplo- pes. *Immunogenetics 27:* 110–120, 1988

gall, K. T., Wright, H., Redmond, J., Dutia, B. M., Hopkins, J., ng, J., Deverson, E. V., Howard, J. C., Puri, N., and Haig, D. M. xpression and characterisation of ovine major histocompatibility mplex class II (OLA-DR) genes. *Anim Genet 23:* 347–359, 1992

bell, R. D. and Trowsdale, J. A map of the human MHC. *mmunol Today 14:* 347–352, 1993

rson, E. V., Wright, H., Watson, S., Ballingall, K. T., Huskisson, , Diamond, A. G., and Howard, J. C. Class II major histocompat- ility complex genes of the sheep. *Anim Genet 22:* 211–225, 1991

S. A., Maddox, J. F., Gogolin-Ewens, K. J., Baker, L., Wu, M. J., d Brandon, M. R. Isolation, characterisation and evolution of ine MHC class II *DRA* and *DQA* genes. *Anim Genet 24:* 9–255, 1993

Hedges, S. B. The number of replications needed for accurate estimation of the bootstrap P value in phylogenetic studies. *Mol Biol Evol 9:* 366–369, 1992

Jonsson, A.-K. and Rask, L. Human class II *DNA* and *DOB* genes display low sequence variability. *Immunogenetics 29:* 411–413, 1989

Karlsson, L., Suhr, C. D., Sprent, J., and Peterson, P. A. A novel class II MHC molecule with unusual tissue distribution. *Nature 351:* 485–488, 1991

Karlsson, L. and Peterson, P. A. The alpha chain gene of *H-2O* has an unexpected location in the major histocompatibity complex. *J Exp Med 176:* 477–483, 1992

Kimura, M. *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge, 1983

Lenstra, J. A., van Boxtel, J. A. F., Zwaagstra, K. A., and Schwerin, M. Short interspersed nuclear element (SINE) of the Bovidae. *Anim Genet 24:* 33–39 (1993)

Rosen-Bronson, S. and Long, E. O. An unusual form of alternative splicing in the *HLA-DNA* gene. *Immunogenetics 33:* 124–131, 1991

Saitou, N. and Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol 4:* 406–425, 1987

Scott, P. C., Choi, C.-L., and Brandon, M. R. Genetic organisation of the ovine MHC class II region. *Immunogenetics 25:* 116–122, 1987

Slade, R. W., Hale, P. T., Francis, D. I., Marshall Graves, J. A., and Sturm, R. A. The marsupial MHC: the tammar wallaby, *Macropus eugenii*, contains an expressed *DNA*-like gene on chromosome 1. *J Mol Evol 38:* 496–505, 1994

Tonnelle, C., Demars, R., and Long, E. O. DOB: a new *B* chain gene with a distinct regulation of expression. *EMBO J 4:* 2839–2847, 1985

Trowsdale, J. and Kelly, A. P. The human class II *A* chain gene *DZA* is distinct from genes in the *DP, DQ* and *DR* sub-regions. *EMBO J 4:* 2231–2237, 1985

Trowsdale, J., Groves, V., and Arnson, A. Limited MHC polymorphism in whales. *Immunogenetics 29:* 19–24, 1989

Wickens, M. How the messenger got its tail: addition of poly(A) in the nucleus. *Trends Biochem Sci 15:* 277–281, 1990

Wright, H., Ballingall, K. T., and Redmond, J. The *DY* sub-region of the sheep MHC contains an *A/B* gene pair. *Immunogenetics 40:* 230–234, 1994

Young, J. A. T. and Trowsdale, J. The *HLA-DNA* (*DZA*) gene is correctly expressed as a 1.1 kb mature mRNA transcript. *Immunogenetics 31:* 386–388, 1990

Immunogenetics (1995) 43: 76–79

BRIEF COMMUNICATION

Harry Wright · James Redmond · Keith T. Ballingall

The sheep orthologue of the *HLA-DOB* gene

Received: 29 June 1995 / 18 August 1995

The class II region of the sheep major histocompatibility complex (MHC; Ovar-*Mhc*) has been shown to contain orthologues of many of the human MHC (*HLA*) class II genes with the exception of the *HLA-DP* genes. Genes of the *Ovar-DQ* (Scott et al. 1991a, b; Fabb et al. 1993; Wright and Ballingall 1994; van Oorschot et al. 1994) and the *Ovar-DR* (Scott et al. 1991b; Ballingall et al. 1992; Fabb et al. 1993; Schwaiger et al. 1993) subregions have been cloned and sequenced. Recently, the sheep orthologue of the *HLA-DNA* gene has also been described (Wright et al. 1995). The *B* gene partner for the *HLA-DNA* gene (Trowsdale and Kelly 1985) is the non-polymorphic *HLA-DOB* gene (Tonnelle et al. 1985; Servenius et al. 1987) which is situated 160 kilobases (kb) telomeric to the *HLA-DNA* gene (Campbell and Trowsdale 1993). Although cDNA clones of the *HLA-DOB* gene have been isolated, transcriptional levels are very low in B-cell lines and absent in IFN-γ-induced fibroblastic cells (Tonnelle et al. 1985; Jonsson and Rask 1989). In the mouse, the equivalent genes *H-2OA* and *H-2OB* (formerly Aβ2, Larhammar et al. 1985) express an MHC class II protein of unknown function, whose tissue distribution is restricted to B cells and epithelial cells of the thymic medulla (Karlsson et al. 1991; Karlsson and Peterson 1992). A rabbit *DOB* cDNA clone has been isolated from spleen and transcripts detected in various tissues including peripheral blood B cells (Chouchane et al. 1993). On the other hand, Stone and Muggli-Cockett (1993) were unable to detect *DOB* transcripts in RNA from cattle peripheral blood lymphocytes using a heterologous *HLA-DOB* cDNA probe.

The presence of a *DOB*-like gene was inferred from Southern blots of genomic DNA from sheep (Scott et al. 1987) and cattle (Andersson and Rask 1988) using an *HLA-DOB* probe. However, the only *DOB*-like sequences in the nucleotide databases are those of the human (Tonnelle et al. 1985), mouse (Larhammar et al. 1985), rabbit (Chouchane et al. 1993), and chimpanzee (Kasahara et al. 1989). The strong sequence conservation between orthologous genes from evolutionary disparate groups argues that these genes encode functional products. To extend the study of the *DN/DO* region to another member of the ungulate family, we report here the cloning, sequencing, and expression of the sheep MHC class II *DOB* gene.

A clone which hybridized weakly to *Ovar-DQB* and *Ovar-DRB* gene probes was obtained from a cosmid library. The clone was mapped using the enzymes *Eco* RI, *Bam* HI, *Hin* dIII, *Sac* I, and *Sma* I and shown to contain a single *B* gene. The nucleotide sequence of the region which contained exons 1, 2, and 3 is given in Figure 1 together with the predicted amino acid translations. The amino acid translations of exons 1–3 are compared with those of the human, mouse, and rabbit *DOB* genes in Figure 2 (the translation of the chimpanzee cDNA differed from the human gene at only three positions). The amino acid identities between the sheep gene and the *HLA-DOB*, *H-2OB*, and *Orcu-DOB* at exon 1 were 62%, 42%, and 62% respectively. The corresponding figures for exon 2 were 80%, 76%, and 87%, and for exon 3 were 81%, 76%, and 70%. The sheep and human sequences contained the same variation, TTCCAATCC, on the consensus eucaryotic CAAT box, GG(C or T)CAATCT, 120 base pairs (bp) upstream of the initiation codon. There is a putative glycosylation site NGT at amino acid position 20 and the sequence has cysteine residues at positions 15, 79, 117, and 173 conserved in all class II β chains. Exons 4–6 of the human *DOB* gene occur within 1.1 kb of the end of exon 3. We were unable to recognize any comparable open reading frames in the same region in the sheep, nor did the full-length *HLA-DOB* cDNA probe hybridize to a further 0.7 kb downstream of the sequenced region reported here (data not shown). The last 400 bp of the reported sequence are 82%

The nucleotide sequence data reported in this paper have been submitted to the EMBL nucleotide sequence database and have been assigned the accession numbers z49879 and z49880

H. Wright (✉) · J. Redmond · K. T. Ballingall[1]
Moredun Research Institute, 408 Gilmerton Road, Edinburgh, EH17 7JK, UK

Present address:
[1] ILRAD, PO Box 30709, Nairobi, Kenya

```
                                Z box                    X1 box                          Y box
ctgcaggtgaaaaattctggagtggtatcacatgactttatgaaggcatcctacccagcacctaatgacactaagctcctctagcttgattggttatctggatgaggagtttccaatcccgggggt 125
                                                                                                                  exon 1
                                                                                                                M  S  P  I  W  V  P
agacagaaagacatctctgctctgcatgaatatttgactgagtgtgtgtggccaagcccattctaaagaggtggagatttcgtgtcttgtttgctttcttccagaATGAGTCCTATCTGGGTTCC 250

  W  V  V  A  F  S  S  T  V  L  R  L  D  A  S  M  T  Q  G  R  D  S  P
CTGGGTGGTGGCCTTCTCATCCACGGTTCTCAGGCTGGATGCTTCTATGACTCAAGGCAGAGATTCTCCAGgtaagaagatagtgatttctttcaggtatagaggcacacattggcacgggggag 375
agaggcctttccctacaggtccaggttccagtccagacaggaagcagctctgacttcagagaatcttttctcagagaaaagctt.....1.5 kbp not sequenced..... 458
atggacagacaccacctaagcgactgaactgcatattccattgtatatgtgtaccacttttttttttttttaaaagaatagatgtctgtttgctgagaggaaccagggtgtggcccagatgagcag 125
aggctactttgtcagatttactgaggtcaaagccctcctctccaatccatatggtgcattgtaatataccaggtaggggtttcctaaagatgggggctggttcatagtctttttggttttccag 250
exon 2
E  D  F  V  T  Q  A  K  A  D  C  Y  F  T  N  G  T  E  K  V  R  F  V  V  R  F  I  F  N  L  E  E  Y  A  R  F  D  S  D  L  G  M
AAGATTTTGTGACCCAGGCAAAGGCTGACTGTTACTTCACCAATGGGACAGAAAAAGTGCGGTTTGTGGTCAGATTCATCTTCAACCTGGAGGAGTATGCACGTTTCGACAGCGACTTGGGAATG 375

  F  V  A  L  T  E  L  G  K  P  D  A  E  L  W  N  N  R  P  D  I  L  A  R  S  R  A  S  V  D  M  L  C  R  R  N  Y  K  L  G  A  P
TTTGTGGCCTTGACGGAGCTGGGGAAGCCCGATGCTGAGCTGTGGAACAATCGGCCGGATATTCTGGCGAGGAGTAGAGCCTCTGTGGACATGCTTTGCAGACCCAACTACAAGCTGGGTGCACC 500

  F  T  V  G  R  R
CTTCACCGTGGGGAGGAGAGgtgaggtgggagttggggtctcctaggtgggagcctgggtgtggctcttcttccttgtccgggcagtttcctgcttcaggatatctatccatcagggtcaggtgc 625
aggggagactatatatgagctgtccccagttgtgactcactctcactgtcctcaggggtcaggtcactgtgatcttggtcccagataccagttcttggcggccctttgagggctagtgtcttcctg 750
aacacaaggaaaattgtaggtgtgtctgataaagcccacatcctccttgtggacactgacatttccagtgtctgagtcctgggggtgagctttccccacatttccggttgccaggtgggtgatct 875
                                                                                                  exon 3
                                                                                                V  Q  P  E  V  T  V  Y  P  E
ggggaagagactcagggcagagacagccccctaaaggcactccggtgttggctttgtggatttccttcaccacctctcatttttttctctcctcctagTGCAACCAGAGGTGACAGTGTATCCGGAG 1000

  K  T  P  A  L  Q  H  R  N  L  L  L  C  L  V  T  G  F  Y  P  G  D  I  K  V  T  W  F  R  N  G  Q  E  Q  R  E  G  V  M  S  T  G
AAGACCCCAGCCCTGCAGCACCGCAATCTGCTGCTTTGCTTGGTGACAGGTTTCTACCCAGGGGACATCAAGGTCACCTGGTTCCGGAATGGGCAGGAACAAAGAGAGGGGGTCATGTCCACCGG 1125

  L  I  R  N  G  D  W  T  F  Q  T  T  V  M  L  P  M  T  P  E  L  G  E  V  Y  T  C  L  V  D  H  F  S  L  P  S  P  V  S  V  E
CCTCATCAGGAATGGAGACTGGACCTTTCAGACGACGGTGATGCTGCCCATGACGCCTGAGCTTGGAGAGGTCTACACCTGCCTTGTTGACCATCCCAGCTTGCCGAGCCCTGTTTCTGTGGAGT 1250

W
GGAgtgagaaccagcttgtacttctgaacccaacttgggtcgatctgaccctggacagggcctgtgttaccttccgtattcaccttggccacagtactagttcctccctctgacctgaaggagtca 1375
tagaaattgggttcgttgggctgggttggagaaaatcatggcagatagctactcttggatcacatgtatgtcaaaacatgcactgcttggctcgataaatagaaaattactaatacaaatgtacaaag 1500
aatccttcactgaaattaatgcttcttgagtttactccaagttatatcacatcttgttttcagttctgttgaacacattgtgactaccatcccaccactggaaatcatgatcatgatttccagct 1625
atggttcctttgaaggggctcagatctgagaactgcgagcaccgtggacagcacctggggtcactctaactcccctagagggaggcgagggcagatgctgacaaggacctcagggtgaatgagaaa 1750
gaagtcatagcaggagacaggttctcatgaatgtgctgtaaatacaccatgacttttttttttgcaattagagagaatgcagtaaatgcagtctttaaatatagtcaaagctatggttttttca 1875
atagtcatgtacaggtgtgagagctgggccataaagaacgctgagcactaaagactgatgctttcaaatttgcagtgcaggagaagactcttcagggtccctgaactgcaagaaaattaaaccag 2000
tcaatccgaaaagaaaccaaccctgatattcattggaaggactgatgctgaagctgaagctccgatacttttggtcacttgacatgaagagctgactcactggaaaagaccctgatgctgggaaag 2125
accaagggcaaaggagaaggggcagcagaggatgagatggttggatagcactgccaactcaatggacaccaatttgagcaagctctggcagacagggaaggacaggggagcctggcatgctgca 2250
```

Fig. 1 The nucleotide sequence and predicted amino acid translations of exons 1–3 of the sheep *DOB* gene. Exons are in *uppercase* with the splice junctions *underlined*. The core Z, X_1, and Y boxes are *double underlined*. The *caat* signal and the conserved glycosylation site in exon 2 are in *bold* type

similar to the Bov-2 SINE sequence reported by Lenstra and co-workers (1993).

Either intron 3 in the sheep gene is unusually long or the sheep *DOB* gene has suffered a deletion. A third possibility is that a spurious ligation has occurred during construction of the cosmid library. Intron 2 of the sheep gene is 61% similar to that of the human gene over its entire length. On the other hand, while the first 190 bp of intron 3 from the sheep and human *DOB* genes share 66% similarity, the rest of the sequence is not significantly similar. Furthermore, an *Mbo* I site, GATC, occurs at position 1433 in Figure 1, the point at which the similarity in intron 3 breaks down. *Mbo* I was the enzyme used to partially digest the genomic DNA for the library construction. To check for the presence of a cloning artefact, genomic DNA and the cloned DNA were digested with *Eco* RI, *Bam* HI, *Hin* dIII, and *Sac* I and Southern blots hybridized to an exon 2-specific probe from the sheep *DOB* gene. *Hin* dIII and *Sac* I fragments of 9 kb and 5 kb were present in the genomic blot, which were not present in the cosmid. It would appear therefore that cosmid 305 is a chimaera and exon 4–6 have simply not been cloned.

In view of the cloning artefact present in cosmid 305, it was not surprising that we were unable to detect expression of an *Ovar-DN* protein following co-transfection of cosmids containing the *Ovar-DNA* gene and the *Ovar-DOB* gene into mouse L cells. However, we were also unable to detect sheep *DOB* transcripts using a homologous probe on northern blots of RNA from various tissues including spleen, lymph node, and thymus, which were positive for the *Ovar-DRB* isotype (data not shown). Stone and Muggli-Cockett (1993) were unable to detect *Bota-DOB* RNA in peripheral lymphocytes using the heterologous human *DOB*

Fig. 2 Multiple alignment of the amino acid translations of exons 1, 2, and 3 of the human, sheep, mouse, and rabbit *DOB* genes. *Dashes* indicate identity, *asterisks* indicate deletions

```
                Signal peptide
HLA-DOB         MGSGWVPWVVALLVNLTRLDSSMTQG
Ovar-DOB        -SPI-------FSSTVL---A-----
H-2OB           --A-RA----------M----F-IE-
Orcu-DOB        -D-K-----------I---Y-----

                β1 Domain
                1         10        20        30        40        50        60        70        80        90
HLA-DOB         TDSPEDFVIQAKADCYFTNGTEKVQFVVRFIFNLEEYVRFDSDVGMFVALTKLGQPDAEQWNSRLDLLERSRQAVDGVCRHNYRLGAPFTVGRK
Ovar-DOB        R-------T---------------R------------A-----L-------E--K----L--N-P-I-A---AS--ML--R--K---------R
H-2OB           R---N*-----------------HLL----------L-----L----------E---D---K------T--A--NM---QK-K-------E-N
Orcu-DOB        *---------------------Q-RV----------AH----V-------E-------S--N-P-I-----RS--FL--R-------------

                β2 Domain
                100       110       120       130       140       150       160       170       180
HLA-DOB         VQPEVTVYPERTPL*LHQHNLLHCSVTGFYPGDIKIKWFLNGQEERAGVMSTGPIRNGDWTFQTVVMLEMTPELGHVYTCLVDHSSLLSPVSVEW
Ovar-DOB        ----------K--A*-QHR---L-L----------VT--R----Q-E------L--------------P------E--------P--P-------
H-2OB           *P-----------*-Q-----L----------SV---R------S------LV-------G-T-----I----DI-S---E-PG--R----A-
Orcu-DOB        -P--G--I-----VCGQ-PAA-L-D**-LLS---RVR-LR--------I----L----------TE---------D--------P---------Y
```

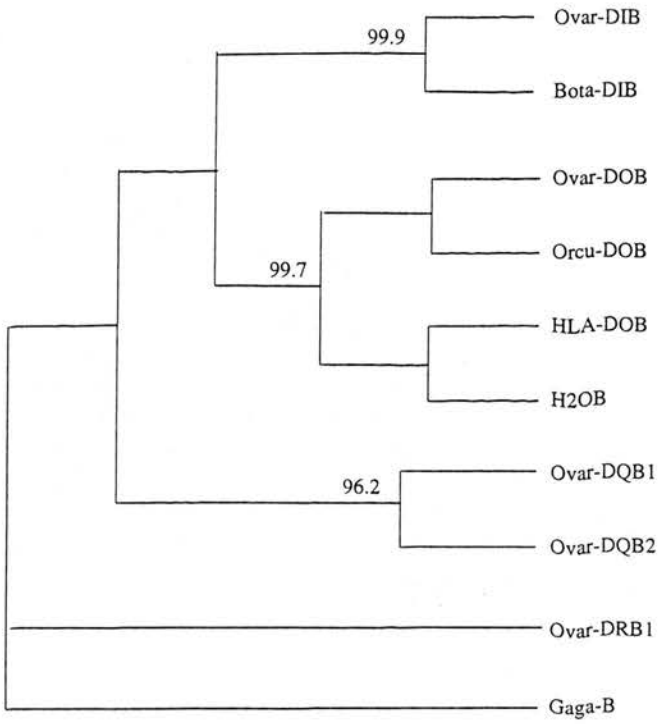


Fig. 3 Phylogenetic tree of selected MHC class II *B* genes. Tree construction is described in the text.
HLA Human (*DOB*, acc. no. m26040);
H2 Mouse (*OB*, acc. no. m11800);
Orcu Rabbit (*DOB*, acc. no. m96942);
Bota Cattle (*DIB*, Stone and Muggli-Cockett 1990);
Ovar Sheep (*DIB*, acc. no. z27400; *DQB1*, acc. no. z28422; *DQB2*, acc. no. z28425; *DRB1*, z11520);
Gaga Chicken, (*B* gene, acc. no. m26306)

gene as a probe. The amount of *DOB* RNA may have been below detectable level. These results contrast markedly with those of Chouchane and co-workers (1993) who found rabbit *DOB* transcripts in spleen, lymph nodes, and appendix. The human (Tonnelle et al. 1985) and chimpanzee (Kashahara et al. 1989) *DOB* cDNAs were obtained from Epstein-Barr virus (EBV)-transformed B cell lines but the level of *HLA-DOB* RNA was more than thirty times lower than that of the *HLA-DRB* isotype. Sequence motifs within the proximal promoter region play crucial roles in the regulation of class II transcription (reviewed by Benoist and Mathis 1990; Glimcher and Kara 1992). Recently, Voliva and co-workers (1993) showed that a construct containing the *HLA-DOB* core Z box of seven nucleotides, in place of the *DRA* Z box, drastically reduced the level of transcription from the *HLA-DRA* promoter in EBV-transformed B cell lines and in IFN-γ-induced fibroblastic cells and attributed this to sequence changes in the core Z box. The sequence of the *Ovar-DRA* Z box is identical to that of the *HLA-DRA* gene. The equivalent sequence in the sheep *DOB* gene is double-underlined in Figure 1. The sheep and human *DOB* Z boxes show three and four nucleotide changes from the *DRA* Z box, respectively. It is therefore likely that a similar mechanism may silence both genes. On the other hand, the core Z box in the *H2-OB* gene shows six changes compared with that of the *HLA-DRA* gene. The equivalent region of the rabbit *DOB* gene has not been sequenced.

A phylogenetic tree of the translated second exons from the four *DOB* sequences together with those from representative sheep and cattle *B* gene sequences is shown in Figure 3. The PHYLIP package was used for phylogenetic analysis. The tree was constructed by applying the neighbor-joining method (Saitou and Nei 1987) to pairwise distances that had been corrected for multiple hits in the amino acid sequence by the method of Kimura (1983). Statistical testing of the tree topology was done using the bootstrap method based on 2000 trials (Hedges 1992). The bootstrap results are summarized as the percentage of trials supporting a given node. Only bootstrap percentages of over 95% are reported. The [(*HLA-DOB*, *H-2OB*) (*Ovar-DOB*, *Orcu-DOB*)] cluster was supported by 99.7% of the trials, suggesting that this cluster was distinct from the rest of the tree. Interestingly, the fork which brought together the *DOB* and the ruminant-specific *DIB* genes (Stone and Muggli-Cockett 1990; Wright et al. 1994) occurred in 1672 of 2000 trees (84%).

In summary, a partial sequence of the *Ovar-DOB* gene is presented, formally demonstrating its presence in the sheep MHC. The sheep gene appears not to be expressed.

Acknowledgments The work was supported by funds from the Scottish Office Agriculture and Fisheries Department (SOAFD). This work benefitted from the use of the SEQNET facility. The clones are available to interested parties.

References

Andersson, L. and Rask, L. Characterisation of the MHC class II region in cattle. The number of *DQ* gene varies between haplotypes. *Immunogenetics 27:* 110–120, 1988

Ballingall, K. T., Wright, H., Redmond, J., Dutia, B. M., Hopkins, J., Lang, J., Deverson, E. V., Howard, J. C., Puri, N., and Haig, D. M. Expression and characterisation of ovine major histocompatibility complex class II (OLA-DR) genes. *Anim Genet 23:* 347–359, 1992

Benoist, C. and Mathis, D. Regulation of major histocompatibility complex genes: X, Y and other letters of the alphabet. *Annu Rev Immunol 8:* 681–715, 1990

Campbell, R. D. and Trowsdale, J. A map of the human MHC. *Immunol Today 14:* 347–352, 1993

Chouchane L., Brown, T., and Kindt T. J. Structure and expression of a transcribed rabbit class II gene with homology to *HLA-DOB*. *Immunogenetics 38:* 64–66, 1993

Fabb, S. A., Maddox, J. F., Gogolin-Ewens, K. J., Baker, L., Wu, M. J., and Brandon, M. R. Isolation, characterisation and evolution of ovine MHC class II *DRA* and *DQA* genes. *Anim Genet 24:* 249–255, 1993

Glimcher, L. and Kara, C. J. Sequences and factors: a guide to MHC class II transcription. *Annu Rev Immunol 10:* 13–49, 1992

Hedges, S. B. The number of replications needed for accurate estimation of the bootstrap P value in phylogenetic studies. *Mol Biol Evol 9:* 366–369, 1992

Jonsson, A.-K. and Rask, L. Human class II *DNA* and *DOB* genes display low sequence variability. *Immunogenetics 29:* 411–413, 1989

Karlsson, L., Suhr, C. D., Sprent, J., and Peterson, P. A. A novel class II MHC molecule with unusual tissue distribution. *Nature 351:* 485–488, 1991

arlsson, L. and Peterson, P. A. The alpha chain gene of *H-2O* has an unexpected location in the major histocompatibility complex. *J Exp Med 176:* 477–483, 1992

asahara, M., Klein, D., and Klein, J. Nucleotide sequence of a chimpanzee *DOB* cDNA clone. *Immunogenetics 30:* 66–68, 1989

imura, M. *The Neutral Theory of Molecular Evolution,* Cambridge University Press, Cambridge, 1983

arhammar, D., Hammerling, U., Rask, L., and Peterson, P. A. Sequence of gene and cDNA encoding murine MHC class II gene Aβ2. *J Biol Chem 260:* 14111–14119, 1985

enstra, J. A., Zwaagtra, K. A., Van Boxtel, H. A. F., and Schwerin, M. SINE sequence in domestic artiodactyles. *Anim Genet 24:* 33–39, 1993

aitou, N. and Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol 4:* 406–425, 1987

chwaiger, F. W., Weyers, E., Epplen, C. Brun, J., Ruff, G., Crawford, A., and Epplen, J. T. The paradox of MHC-DRB exon/intron evolution: alpha-helix and beta-sheet encoding regions diverge while hypervariable intronic simple repeats co-evolve with beta-sheet codons. *J Mol Evol 3:* 260–272, 1993

cott, P. C., Choi, C.-L., and Brandon, M. R. Genetic organization of the ovine MHC class II region. *Immunogenetics 25:* 116–122, 1987

cott, P. C., Gogolin-Ewens K. J., Adams T. E., and Brandon M. R. Nucleotide sequence, polymorphism, and evolution of ovine MHC class II *DQA* genes. *Immunogenetics 34:* 69–79, 1991 a

cott, P. C., Maddox, J. F., Gogolin-Ewens, K. J., and Brandon, M. R. The nucleotide sequence and evolution of ovine MHC class II *B* genes: *DQB* and *DRB*. *Immunogenetics 34:* 80–87, 1991 b

Servenius, B., Rask, L., and Peterson, P. A. Class II genes of the human MHC: the DOβ gene is a divergent member of the class II β gene family. *J Biol Chem 262:* 8759–8766, 1987

Stone, R. T. and Muggli-Cockett, N. E. Partial nucleotide sequence of a novel bovine major histocompatibility complex class II beta chain gene, *BoLA DIB*. *Anim Genet 21:* 352–360, 1990

Stone, R. T. and Muggli-Cockett, N. E. *BoLA-DIB:* species distribution, linkage with *DOB* and northern analysis. *Anim Genet 24:* 41–45, 1993

Tonnelle, C., Demars, R., and Long, E. O. *DOB*; a new *B* chain gene with a distinct regulation of expression. *EMBO J 4:* 2839–2847, 1985

Trowsdale, J. and Kelly, A. P. The human class II *A* chain gene *DZA* is distinct from genes in the *DP, DQ,* and *DR* sub-regions. *EMBO J 4:* 2231–2237, 1985

Van Oorschot, R. A. H., Maddox, J. F., Adams, L. J., and Fabb, S. A. Characterisation and evolution of ovine MHC class II *DQB* sequence polymorphism. *Anim Genet 25:* 417–424, 1994

Voliva, C. F., Tsang, S., and Peterlin, B. M. Mapping *cis*-acting defects in promoters of transcriptionally silent *DQA2, DQB2* and *DOB* genes. *PNAS USA 90:* 3408–3412, 1993

Wright, H. and Ballingall K. T. Mapping and characterisation of the *DQ* subregion of the ovine MHC. *Anim Genet 25:* 243–249, 1994

Wright, H., Redmond, J., and Ballingall, K. T. The *DY* sub-region of the sheep MHC contains an *A/B* gene pair. *Immunogenetics 40:* 230–234, 1994

Wright, H., Redmond, J., Wright, F., and Ballingall, K. T. The nucleotide sequence of the sheep MHC class II *DNA* gene. *Immunogenetics 41:* 131–133, 1995